

Greens can be good for you

Next month's general election in Germany may see the end of the Green Party's spell in government. The party has fared well, as has science with it, except where ideology won out over good sense.

The Green movement in Germany was born almost 30 years ago when a loose coalition of pacifists, antinuclear campaigners and socialist intellectuals united to form an extra-parliamentary movement of discontent and nonconformism. The surge quickly gained momentum and became a political party that sought to pursue the cause of the environment and little else.

With hindsight, the party's romantic attachment to the ideals espoused by philosopher Jean-Jacques Rousseau in the eighteenth century was absurdly anachronistic, opposing almost everything that doesn't grow on trees — from artificial fertilizers to plastic bags and computers. No wonder that the establishment, and many scientists, were deeply suspicious of the emerging new force.

Over the years, the Greens have jettisoned their excess anti-modernism. Nevertheless, when the reformed party came to power in 1998 as the Social Democrat's coalition partner, scientists in Germany were nervous about where it might lead. Now, with an upcoming election in which the Greens are unlikely to retain power, it is an appropriate moment to take stock of their achievements.

Welcome progress

By and large, and with some notable exceptions, the Greens have helped sustain and develop Germany's science base. Public science budgets have increased during the past seven years, which they hadn't under the previous government. In areas where the Green's core interests reside, research opportunities have flourished. Encompassing such essential disciplines as alternative and renewable energies, clean environment technologies, biodiversity, coastal protection and marine sciences, this is welcome progress.

Despite early fears, the Greens have not fuelled controversy over the use of animals in research, nor have they been responsible for Germany's overly restrictive regulation of stem-cell research. Indeed, the loudest opposition to some aspects of modern biomedicine comes from the conservative Christian Democrats, who look likely to win the election if their leader Angela Merkel can quickly improve on her shaky initial campaign.

The Green ministers for agriculture and the environment, Renate Künast and Jürgen Trittin, respectively, have gained Europe-wide respect for their political handling of two key issues: Künast for her rigorous crisis management of the BSE (mad cow disease) calamity, and Trittin for enticing German industry into an agreement on greenhouse-gas emissions trading. Green power and science have also prospered alongside each other at a municipal level. Witness Konstanz and Freiburg, for example — cities in southwest Germany which both host large research universities and, thanks to the large student electorate, are governed by Green mayors. But it is mostly thanks to Joschka Fischer, the popular Green minister of foreign affairs, that the former single-issue party is now respected for its broad competence in societal issues.

Blame where it's due, however: Green politicians, and their voters, still display deep-rooted reservations when it comes to technological advance as a means of solving problems. This is where Green principles have clashed with those of the Social Democrat science minister, Edelgard Bulmahn.

Tensions escalated in March this year over a grant application to the research ministry for risk research on transgenic plants. The application came from the Federal Biological Research Centre for Agriculture and Forestry, which is subordinate to Bulmahn's ministry. Künast insisted that the application should be withdrawn. The reasons were never clear, and Künast has not shaken off the suspicion that she blocked research to which she was hostile.

Hostilities have been most manifest in the Green's approach to nuclear energy. Phasing out nuclear energy by 2020 was a cornerstone of the Social-two parties' coalition agreement. But while it was in accord with the public mood, it also inevitably led to an unprecedented loss of scientific competence in fission-reactor technology and safety research, and was an outrageous waste of capital. Unsurprisingly, nuclear-physics departments at German universities are already finding it difficult to attract students and PhD candidates.

"Greens still display deep-rooted reservations when it comes to technological advance as a means of solving problems."

Missed opportunities

It is also thanks to the Greens that Germany will contribute less than it should to the construction of the international fusion reactor ITER in Caderache, France. Sceptical Green members of the parliamentary research committee have destroyed a substantial opportunity for German scientists and engineers.

The same applies to plant biotechnology. Although Germany is well positioned in plant genetics, the restrictive regulatory framework set up by Künast makes it difficult to reap the benefits of basic research. Recently proposed liability rules are far stricter than anywhere else in the European Union and will deter investors. If the law passes, Germany could suffer the same fate in its crop trade as it did in the past with genetically engineered drugs. For example, despite its strong pharmaceutical industry, Germany has to import insulin.

No one should expect the Greens to abandon their principles, but one can expect them to be more prepared to end short-sighted hostilities in the longer-term interests of the sustainable environment and society about which they care so much.

The German experiment has shown that, such myopias aside, the Green party has been more open-minded, versatile and science friendly than most had believed it capable of. A right-liberal government (or a grand coalition) has yet to prove that it can do better for science and for society's resources. ■



Join a social revolution

Your bookmarks make your web life manageable.
But we can all benefit by sharing them.

Is the big challenge in the Internet era information overload or underload? Those who complain about the former may simply be inadequately organized. Many tools now exist for taming the flow of scientific information on the web, but scientists have been slow to adopt them and are no doubt missing out on gems as a result.

Take, for example, the need to manage the results of a web search. Storing the items selected so that you can easily find them again is often critical. A new generation of 'social bookmarking' services now allow a user to post an article or web page with a single click to a personal web collection and to automatically group them under keyword tags. The 'social' element arises from the fact that these bookmarks can easily be shared over the web, either selectively or publicly.

Three leading examples of such services are del.icio.us (<http://del.icio.us>), CiteULike (www.citeulike.org) and Connotea (www.connotea.org). The last of these is an open-source service produced by *Nature's* publishers, the Nature Publishing Group, and will be used in the following examples, but CiteULike is also tailored for scientists. There is no implication here that one service is better than another. The intention is simply to highlight ways by which readers can make more efficient use of the web and spread their knowledge at the same time.

Users of Connotea say how easy it is to post and tag an article as they are looking at it on, say, Pubmed. And once a user clicks in their browser to send an article to Connotea, the software looks up the various metadata for the article, such as authors, journal volume and page numbers, and adds these to the entry.

Kevin Olbrich, a researcher at Duke University Medical Centre in

Durham, North Carolina, is one enthusiast. A key difference from personal desktop reference software is that many of Olbrich's bookmarks can be viewed by anyone (<http://www.connotea.org/user/kco>), as can those of any user.

Some researchers are reluctant to use such services because they do not want their competitors to see what they are reading. But users can opt to make a bookmark private if they wish. Or they can create a 'group' with other users, and make their bookmarks available only to members of that group.

There is added value here. Whereas a PubMed or Google search will bring back everything that matches a keyword, the fact that a researcher considers an article important enough to post to his or her collection implicitly says something about the likely value of the article. There's no greater sign that an article is important than when a colleague e-mails you to say "Hey, you should see this." Connotea takes the same principle but broadens the group who benefit. Other researchers can be alerted by an RSS feed ([http://en.wikipedia.org/wiki/RSS_\(protocol\)](http://en.wikipedia.org/wiki/RSS_(protocol))).

Such social collections share the effort that people have put into searching. Suppose you are contemplating introducing electronic laboratory notebooks (ELNs) to your laboratory. Search on Google for ELN and you may well find that a Connotea collection is the number one hit. You can thus save effort by using the collection of Connotea links on electronic laboratory notebooks (www.connotea.org/tag/ELNS) discovered by a *Nature* journalist while researching a feature article.

These are early days and suppliers of these services have plenty to learn. But readers are urged to try them out, provide feedback and join the quiet social revolution. ■

"There's no greater sign that an article is important than when a colleague e-mails 'Hey, you should see this.' Connotea takes the same principle but broadens the benefit."

Who owns your work?

A case in the Kansas Supreme Court reflects a lack of clarity in US copyright law.

A decade ago, US universities, infected with the internet mania of the early nineties, convinced themselves that there was a large amount of money to be made in distance education. The online course materials were not subject to patents, which the university typically owns, but to copyright, which traditionally rests with its academics. So a number of schools made attempts to wrest copyright away from the researchers under the doctrine of 'work for hire'.

Under work-for-hire, if you create something for your employer, copyright belongs to the employer. Academics have not traditionally been included in this category because although everyone agrees that they must publish or perish, and that teaching courses is part of the job, the head of a department does not typically request a specific paper or lecture on a specific subject. "Have a 20-page review of

isopod parasites on my desk by Monday" is not the order of the day.

Besides, if the university owns the scholarly output of its faculty, it is also responsible for that output, and might be tempted to shape it, for example when threats of litigation or bad PR loom.

Almost all universities who tried to claim copyright have backed off. In those cases where work-for-hire seems to be particularly applicable — when a faculty member is asked to produce a specific document, for instance — an agreement is often signed waiving the university's right to the copyright.

But an exception in Kansas (see page 1072) has highlighted the uncertainties over ownership of intellectual property in current US law. If the Kansas Supreme Court rules that work-for-hire should apply to academics, this will muddy the waters more. If it does not, it will hardly clarify matters. Tradition hands copyright to academics in most cases, but the law is unclear. The time has come to tidy it up. ■

"Tradition hands copyright to academics in most cases, but the law is unclear. The time has come to tidy it up."

RESEARCH HIGHLIGHTS

Sea change

Biol. Lett. doi:10.1098/rsbl.2005.0351 (2005)
Humpback whales (*Megaptera novaeangliae*, pictured) are known for their yearly migrations between the poles and tropics. Calves learn a route from their mothers and follow it every year. And adherence to these routes maintains lineages with distinct genetics and song type. But speculation that a few very bold, or badly lost, whales switch routes has been reinforced by Cristina Pomilla and Howard Rosenbaum, both of the American Museum of Natural History and the Wildlife Conservation Society in New York. One male whale whose DNA was sampled when it was found wintering in the Indian Ocean by Madagascar in 2000 was sampled again in 2002 in the South Atlantic, near Gabon. Microsatellite analysis made the match, and snapshots of the whale's dorsal fin confirmed it.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

A. NACHOUM/CORBIS

CELL BIOLOGY

Calorie burner

Cell Metab. 2, 105–117 (2005)

A role in glucose metabolism in mammals has been revealed for a class of proteins associated with longevity in worms, flies and yeast.

Shin-ichiro Imai and his colleagues of the Washington University School of Medicine, Missouri, engineered mice to overexpress Sirt1 proteins in their pancreatic beta cells. Compared with control mice, the engineered animals produced more insulin, the protein that regulates carbohydrate metabolism, in response to doses of glucose.

Although Sirt1 may be involved in prolonging the life of mammals on calorie-restricted diets, a link has not been proven. The more immediate implication of this study is the possibility that Sirt1 could be used to treat type 2 diabetes, which strikes when the

body becomes resistant to insulin, or when the pancreatic cells produce too little of it.

CANCER

Timing is everything

Cancer Cell 8, 99–110 (2005)

Timing could be the key to successful combination cancer treatments, reports Mark Dewhirst's lab at Duke University Medical Center in North Carolina.

Last year, Dewhirst and his team showed that radiation increases the activity of the hypoxia-inducible factor-1 (HIF-1) protein in tumours, which in turn affects the responsiveness of the tumours to the treatment. Now they have teased apart the pathways through which HIF-1 works — identifying three that make a tumour more radiosensitive, and one that makes it more resistant to radiation.

They conclude that radiation followed by HIF-1 inhibition would be the most effective cancer treatment, but the strength of the effect varies from tumour to tumour — and probably from patient to patient.

NANOTECHNOLOGY

Mini might

Phys. Rev. B 72, 085416 (2005)

The concept of chains and chain-mail constructed from ring-shaped carbon nanotubes has been explored through molecular dynamics, using a new computational method to study the materials' response to loading.

The calculated tensile strength of such

structures is very impressive, given their low mass density. Chains of interlinking nano-rings could lengthen by more than a third without snapping, whereas the mail could withstand a strain of 25%. And unlike the response of structures made from metal, the deformation would also be totally reversible. Although nano-mail has yet to be fabricated, interlinked carbon nano-rings, where each ring measures a few hundred nanometres in diameter, have been observed in experiments.

STEM CELLS

Easy does it

PLoS Biol. doi:10.1371/journal.pbio.0030283 (2005)

Neural stem cells used to be difficult to grow because they had to be cultured alongside more differentiated cells within floating clusters called neurospheres. But a study now suggests that none of this is necessary. Austin Smith of the University of Edinburgh, UK, and his colleagues show that a combination of fibroblast growth factor 2 and epidermal growth factor encourages isolated neural stem cells to propagate. Using this trick, they derived a pure culture of self-renewing neural stem cells from embryonic stem cells obtained from mice. The neural stem cells were able to differentiate into both neurons and their companion glial cells, astrocytes.

NEUROSCIENCE

Underlying Alzheimer's

Nature Med. doi:10.1038/nm1287 (2005)

The vascular lesions in the brain that are a distinctive, but little understood, feature of

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

D. H. WELLS/CORBIS

Alzheimer's disease have been linked to low expression of the transcription factor GAX.

A team led by Berislav Zlokovic at the University of Rochester Medical Center in New York made this discovery by studying gene expression in endothelial cells from the brains of Alzheimer's patients. GAX is known to regulate the development of the vascular system, and restoring GAX levels in these endothelial cells *in vitro* stimulated the growth of blood vessels. It also enhanced expression of a factor that helps clear the protein plaques typically found in the brains of Alzheimer's patients. Changes in the brains of mice lacking one copy of the gene that encodes GAX, *Meox2*, provide supporting evidence of GAX's role.

MICROBIOLOGY

Close encounters

Science **309**, 1245–1248 (2005)

Cell-to-cell contact seems to allow certain bacteria to stymie their rivals. David Low's group at the University of California, Santa Barbara, report that the EC93 strain of *Escherichia coli* transfers growth-inhibiting signals when it comes into contact with neighbouring cells of a different strain.

The EC93 strain was isolated from the guts of rats, where it had eliminated all other bacteria. The team identified two proteins that the bacterium releases to inhibit its neighbours' growth, and also found a DNA sequence in the EC93 strain that provides it with immunity against its own secretions. The researchers speculate that the contact-dependent inhibition may involve interactions between tiny tentacles called pili found on the surface of the cells.

OPTICS

Caught behind bars

Opt. Express **13**, 5961–5975 (2005)

What is the best way to cage light? This is a useful feat in building all kinds of optical devices including lasers. But in the past, researchers have had to use trial-and-error to design photonic-crystal cavities that trap light in very small volumes. Now a team at Stanford University in California has devised an equation that does the job in a single computational step.

The inputs to the equation are the desired pattern of the trapped light field, the volume within which the light must be confined and the quality factor, or leakiness, of the cavity. The outputs are instructions for how to arrange the different layers of material that form the photonic crystal.

NEUROBIOLOGY

A taste sensation

Neuron **47**, 593–605 (2005).

To get the most from your meal, you should savour the scent before tucking in. Dana Small of the John B. Pierce Laboratory in New Haven, Connecticut, and her colleagues show that a smell arriving through the nose can stimulate different regions of the brain

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

ADVERTISING ARCHIVES

compared with the same odour delivered through the mouth.

The team used functional magnetic resonance imaging to probe the brain's response to scents delivered into the nostrils or — to simulate odours arriving through the mouth — to the back of the nasal passage. A difference in response was seen for chocolate, but not for three non-food smells. The team suggests that the mechanism distinguishes between the availability and receipt of food.

BACTERIOLOGY

Tactical defence

Cell **122**, 461–472 (2005)

The bacterial pathogen *Salmonella typhimurium* senses when it has become the target of its host's immune system, and takes steps to avoid destruction. Samuel Miller at the University of Washington Medical School in Seattle and his colleagues have elucidated the mechanism.

They deciphered how the PhoQ enzyme bound to the bacterial cell membrane is activated by the positively charged peptides released by the host cell to kill the bacteria. This triggers a cascade of events that affects the expression of more than 200 genes, including some that strengthen the bacterial outer membrane, and so protect the bacteria from attack.

JOURNAL CLUB

Elizabeth Brainerd
Brown University, Providence

A specialist in functional anatomy hopes that advances in stickleback genetics will help solve a mystery about her favourite group of vertebrates.

I have long been fascinated, for no sensible reason at all, by pufferfish and their various spiny, prickly and armoured relatives. Some unknown trait of this group Tetraodontiformes predisposes its members to repeated evolution of mechanical defences against predation, such as body inflation in pufferfish, stout spines in triggerfish and whole-body armour in boxfish.

Impressive body armour and spines are also found in Gasterosteiformes, a group that includes seahorses, pipefish and sticklebacks. Recently, progress has been made in revealing the genetics of armour in the threespined stickleback (*Gasterosteus aculeatus*).

The question tackled by the new research is how freshwater sticklebacks evolved to have fewer armoured plates along their sides than marine sticklebacks. The populations separated after the end of the last glaciation, some 20,000 years ago.

Unexpectedly, Cresko *et al.* showed that the loss of lateral armour plates has the same genetic basis in geographically isolated freshwater populations (*Proc. Natl Acad. Sci. USA* **101**, 6050–6055; 2004). Variation in *Ectodysplasin*, a gene known to affect skin and scale development, was then implicated as the causal factor by Colosimo *et al.* (*Science* **307**, 1928–1933; 2005).

An interesting twist is that the alleles found in low-plated fish also turn up in marine sticklebacks, indicating that evolution in the different freshwater populations was driven by selection on pre-existing alleles, rather than on parallel mutations.

Understanding the evolution of tetraodontiform defences will be a harder problem, in part because they diverged over 50 million years ago. But work such as this gives me hope.

NEWS

P. SULLIVAN/AP

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Painkiller verdict shows mistrust of Merck

Carol Ernst and her attorney win their suit against the drug firm Merck following her husband's death.

A court decision against the maker of Vioxx has set off a wave of legal action.

On 19 August a jury in Angleton, Texas, ordered the pharmaceutical firm Merck to pay \$253 million to Carol Ernst, whose husband died after taking the painkilling drug Vioxx.

The amount is almost certain to be reduced, and the company is likely to challenge the decision to go to trial in the first place based on evidence gathered about Vioxx's side-effects. But other plaintiffs are lining up, with more than 4,000 related cases already in preparation (see *Nature* **436**, 459; 2005).

"This is the first trial of many," says Alabama attorney Andy Birchfield, who will present the claim of another widow against the New Jersey-based company in a federal court this autumn. "It's just a first step, but it does show that plaintiffs can meet the legal burden of showing that Vioxx caused a particular heart attack or stroke."

Robert Ernst died of an irregular heartbeat — arrhythmia — in 2001 after taking the medication for eight months.

Merck initially developed Vioxx (rofecoxib) as a safer alternative to painkillers for arthritis. Older drugs, such as aspirin and ibuprofen, work by inhibiting cyclooxygenase enzymes, but this can sometimes lead to severe gastrointestinal bleeding. Vioxx is more specific in its target: it inhibits cyclooxygenase-2 (COX-2)

without blocking the related enzyme COX-1.

Researchers thought this would avert gut problems, but some experts fear that inhibiting COX-2 alone can also increase the body's ability to produce heart-stopping blood clots.

Several years ago, studies began to emerge suggesting that people taking Vioxx had an increased risk of heart attack. In 2000, Merck started a study to assess the anticancer potential of Vioxx — it hoped the drug would prevent the formation of colon polyps, which typically overexpress COX-2. But it found that participants taking Vioxx for an average of 18 months experienced twice as many cardiovascular events as those on a placebo.

The company ended the trial prematurely and withdrew Vioxx from the market last year.

Plain speaking

Carol Ernst's case hinged as much on the amount Merck knew about Vioxx's risks before the withdrawal as on whether the drug was the direct cause of her husband's death. Jurors said they felt the drug company was hiding information about Vioxx, and voted accordingly.

It was not easy for lawyers to link Robert Ernst's death to the painkiller. Merck supporters point out that there are no studies associating Vioxx with fatal or non-fatal arrhythmias, adding that Robert Ernst had been taking Vioxx for a far shorter time than

the 18 months that was linked to heart attacks.

His autopsy uncovered no evidence of a blood clot. But the former local coroner testified at the trial that a blood clot could have dissipated before the examination.

Jurors had to feel convinced only that it was more likely than not that Vioxx was linked to Robert Ernst's death. And their decision did not have to be unanimous; in fact, two of the twelve jurors sided with Merck. Some observers believe that the outcome had more to do with sending a message to the pharmaceutical company than pinning down the cause of one particular death. Merck's stock fell by almost 8% after the verdict was announced.

Plaintiffs in other countries, including Britain, are now making plans to come to the United States, where legal conditions are more favourable for their claims.

The Ernst case was not considered the strongest among those pending, but the verdict should not give others carte blanche to launch into Merck, warns Howard Erichson of Seton Hall Law School in Newark, New Jersey. "You can't assume just because one plaintiff won that all plaintiffs will win."

Merck is likely to make a series of legal appeals. Benjamin Zipursky of Fordham University Law School in New York says, "There are many ways this verdict can be challenged."

Roxanne Khamsi



VIDEO GAMES CAN MAKE KIDS AGGRESSIVE

Psychological association calls for less violence in games.

www.nature.com/news

Tsunami damage was enhanced by coral theft

Illegal removal of coral along Sri Lanka's coastline increased the amount of destruction wrought on the island by last December's tsunami, say researchers.

Harindra Fernando, a fluid dynamicist at Arizona State University in Tempe, made the connection after a visit to his native Sri Lanka earlier this year. While serving as a scientific expert and translator for a BBC-documentary team, he chatted with locals who said they saw the tsunami deflect sideways when it hit coral — which would have made it less powerful than in coral-free areas. Fernando linked this to trucks he had seen last year carrying mounds of coral away from the sea.

Using the eyewitness reports, estimates of wave heights, and a series of dives to verify the presence or absence of corals, Fernando and his colleagues produced a map of coral gaps and wave inundation along Sri Lanka's southwest coast (H. S. J. Fernando *et al. Eos* 86, 301, 304; 2005).

The tsunami reached significantly farther inland through the gaps: in one instance a jet of water 1.5 kilometres long knocked a passenger train off its tracks, killing 1,700. But only a few kilometres away, where the coral was still intact, the wave travelled just 50 metres inland and caused no deaths.

There is a precedent for this phenomenon. In Nicaragua in 1992, a tsunami surged through a break in the coral reef made to let boats through. "Within this passage, water went one kilometre inland," says Fernando. "But nearby, where the coral was intact, there were still beach umbrellas standing."

In Sri Lanka, coral is illegally mined to provide souvenirs for tourists, or to be ground up

for use in house paint. Coral harvesters sometimes blow the reefs up with dynamite in order to collect fish at the same time. Often, the reefs in the best shape are those in front of hotels, as the hotel owners maintain them for the tourists. Fernando hopes that his finding will encourage the Sri Lankan government to enforce its laws against coral mining.

Policy may very well change, says M. Sanjayan, an ecologist with the Nature Conservancy, who surveyed the environmental damage after the tsunami. "There has been a groundswell of support in Sri Lanka for better enforcement of laws," he says. "There is a window of opportunity right now."

Emma Marris



Barrier reef: coral can help to diminish the power of tsunamis as they approach the coast.

M. SANJAYAN/TNC

Dalai Lama gets go-ahead for meditation lecture

The Dalai Lama will speak at this year's annual meeting of the Society for Neuroscience (SfN) despite a petition calling for the lecture to be cancelled.

Campaigners had collected more than 500 signatures in protest against the talk, which they presented to the society's president, Carol Barnes, on 15 August.

The Dalai Lama had been invited to speak at November's meeting in Washington on the effects of meditation on the brain. But some neuroscientists said that a talk by the Buddhist leader was inappropriate at an academic meeting (see *Nature* 436, 452;

2005). Others accused the Dalai Lama of spreading religious ideas under the guise of scientific research into meditation.

Four days after meeting the protesters, Barnes notified one of the petition's organizers, Bai Lu, a neuroscientist at the US National Institutes of Health, that the lecture would go ahead as planned.

The talk will be the first in a new series of lectures called "dialogues between neuroscience and society". Joe Carey, public information director for the SfN, says that the society's leadership "continues to believe that the original plan and purpose of the

dialogues series makes sense, and that the first two invited speakers are consistent with the intent". The Dalai Lama's talk will be followed by one from architect Frank Gehry at the society's 2006 meeting in New Orleans.

Six abstracts for this year's meeting have been withdrawn by one SfN member in protest against the lecture. But since the controversy became public, the society says that it has received a lot of e-mails on the issue, nearly all of them in favour of the talk.

The president's decision, says Carey, will be the society's final word on the issue.

David Cyranoski

Embezzlement scandal rocks Korean universities

M. HENLEY/PANOS

South Korean universities are to face increased scrutiny after two professors at Seoul National University (SNU) were charged in July with embezzling research funds.

A growing awareness of abuse of research money has spurred a nationwide investigation. Alleged abuses include organizing meetings and entertaining laboratory members with funds that were allocated for different purposes, such as grants for postdocs and other staff.

"Often it's mismanagement rather than embezzlement," says Chan-Mo Park, a digital-image processing specialist and president of Pohang University of Science and Technology on the east coast.

But in the case of Byung-Hwan Oh, a professor of civil engineering at SNU, the money reportedly found its way into his bank account. Oh, who earlier this year won the American Concrete Institute's Wason Medal for Materials Research, has been charged with misappropriating about 1.6 billion won (US\$1.6 million) in research funds.

The charges say that Oh bought research equipment from companies that did not exist and fabricated receipts. Earlier in July, Youngman Cho, a mechanical engineering professor

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Hustle and bustle: will further grant abuse be uncovered in South Korea's scientists?

at SNU, was detained for the misappropriation of 190 million won.

The cases were a shock to "us and to all the citizens", says Myoung-Mo Kim, chair of SNU's civil-engineering department. The cases have led to the resignation of the dean of SNU's college of engineering.

The state prosecutor's office reportedly has evidence against eight more SNU researchers, as well as material on scientists at five other

universities and one national research institute.

Park says the main problem is that SNU and many other Korean universities have no central purchasing office — professors buy their own equipment and research materials. "It's a loophole for unethical people," agrees Sun-Young Kim, a molecular biologist at SNU.

He adds that in South Korea big grant winners get no additional salary for their success, and some feel entitled to skim a bit off the top as a reward.

SNU is now trying to establish a central purchasing system, although some of the university's scientists suspect that it might not be able to afford the major investment in staff that would be needed. But there has to be a change of policy, says Myoung-Mo Kim.

Meanwhile, other universities await a visit from the prosecutor's office. Founded 19 years ago, Pohang University of Science and Technology has had a centralized purchasing system from the beginning, says Park. Nevertheless, he is taking measures to boost morale so that researchers do not feel like cheating.

And he is looking into the university's rules and regulations. "There still could be loopholes. If so, we must close them," he says. ■

David Cyranoski

Kansas to rule on copyright for lecture notes

Should all the writings of faculty scientists automatically belong to their university? The point is being tested in the Kansas Supreme Court, which may rule that all written works produced by faculty members — including lecture notes, books, curricula, websites and articles — are the property of universities by default.

The case began in 1998, when the Kansas Board of Regents, which oversees higher education in the state, produced an intellectual-property policy without consulting faculty members at Kansas's Pittsburg State University. Concerned that researchers could miss out on profits from patents on their inventions, the faculty's union filed a complaint.

When the case arrived at the Topeka-based Court of Appeals in 2004, the complaint was rebuffed. The judge cited the 'work-for-hire' rule in federal copyright law. Under this principle, which holds in many other countries, employees surrender all rights to their employers.

Although work produced by academic researchers is strictly part of their job, it is uncertain whether work-for-hire applies to them, and most universities allow copyright to rest with their employees. "The law at the federal level is not clear at all," says Polk Wagner, a specialist on intellectual-property law at the University of Pennsylvania in Philadelphia.

It makes sense for employers at software companies or advertising agencies to own the written material produced by their employees, but extending that principle to academia spells disaster, according to Ann Springer, associate counsel at the Washington-based American Association of University Professors.

The association filed voluntary testimony with the Kansas Supreme Court, stating that applying the work-for-hire rule to faculty members "would wreak havoc with settled academic practices". It added: "Academic freedom requires that faculty be free to produce work reflecting their own views

and theories — not those of the university's administration."

After all, says Springer, if a university owns the copyright on a contentious work, it could alter or suppress it. "Scholarly work has to be under the control of the faculty, or it has no value," she says.

However, David Schauner, general counsel for the Topeka-based Kansas National Education Association, which runs the Pittsburg union, says he does not care whether the faculty members are treated as work-for-hire or not. "I think we probably are work-for-hire employees, but that doesn't mean the university can't pay to incentivize my work," he says. "I want a share — this is a capitalist society after all."

Most observers expect the Kansas Supreme Court to overturn the lower court's decision when it rules on 8 September. But if it upholds the work-for-hire ruling, it could influence future cases in other states. ■

Emma Marris

Adult suicides linked to popular antidepressant

A drug that has been associated with increased suicide risk in children may pose the same danger to adults.

A study in the journal *BMC Medicine* has analysed unpublished data on paroxetine, a drug that is used to treat millions of depressed people in Britain every year. Patients taking paroxetine are more likely to attempt suicide than those taking a placebo, according to researchers based at the University of Oslo in Norway.

Paroxetine, which is sold as Seroxat and Paxil, belongs to a class of drugs called selective serotonin re-uptake inhibitors (SSRIs). These drugs increase the availability in the brain of serotonin, a chemical linked to mood and emotions.

In October 2004, the US Food and Drug Administration insisted that labels carry warnings about the use of paroxetine in children. This followed analysis of clinical trials that showed an increase in suicidal thoughts and behaviours in children taking the drug.

Britain's Committee on Safety of Medicines had issued warnings much earlier, in June and September 2003. They advised practitioners that paroxetine should not be used to treat depressive illness in children and adolescents under the age of 18 years. A few months later, in December 2003, the committee extended this warning to cover all SSRIs except fluoxetine (Prozac) for the treatment of depression in children.

But both the UK and US agencies concluded that the data did not confirm an increased risk in adults. They said that the benefits of taking these drugs outweighed the risks for depressed people.

In February 2005, researchers published studies on drug-company data that were submitted for safety review to the London-based Medicines and Healthcare products Regulatory Agency. They found no evidence that SSRIs increase the risk of suicide in adults, but warned doctors that such dangers could not be ruled out (D. Gunnell, J. Saperia and D. Ashby *Br. Med. J.* **330**, 385–388).

Their analysis of paroxetine was incomplete because the data supplied by the manufacturer, GlaxoSmithKline, did not allow researchers to distinguish between suicidal thoughts and suicide attempts.

Ivar Aursnes and his colleagues applied to

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

The drug paroxetine may exacerbate suicidal thoughts in adults as well as children.

the Norwegian Medicines Agency to obtain the unpublished data on paroxetine. After a year of correspondence, they were finally granted access to the original drug-company data. Aursnes and his team looked at 16 studies of depressed adults who had been randomly assigned paroxetine or a placebo.

The researchers compared the number of suicides and suicide attempts in the 916 patients taking paroxetine with that in 550 taking the placebo. Eight patients attempted suicide — seven of them were taking paroxetine. The probability that the drug increased the risk of suicide was 90% in the participants that were studied.

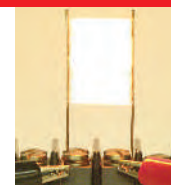
“The use of these drugs needs to be more restricted,” Aursnes told *Nature*. “People should be made aware of the dangers.”

Doctors also prescribe paroxetine for problems such as generalized anxiety disorder and social anxiety. “But the risk of suicide is specific to patients receiving paroxetine for depression,” says Aursnes, because such people may already have suicidal tendencies.

Tim Kendall, co-director of Britain's National Collaborating Centre for Mental Health, says that the results need to be compared with previously published studies. “We need to look at this more carefully,” he says. ■

Jennifer Wild

M. PRINCE/CORBIS



NANOTUBE SHEETS COME OF AGE

Clear, conductive sheets can be produced at high speed

www.nature.com/news

SNAPSHOT

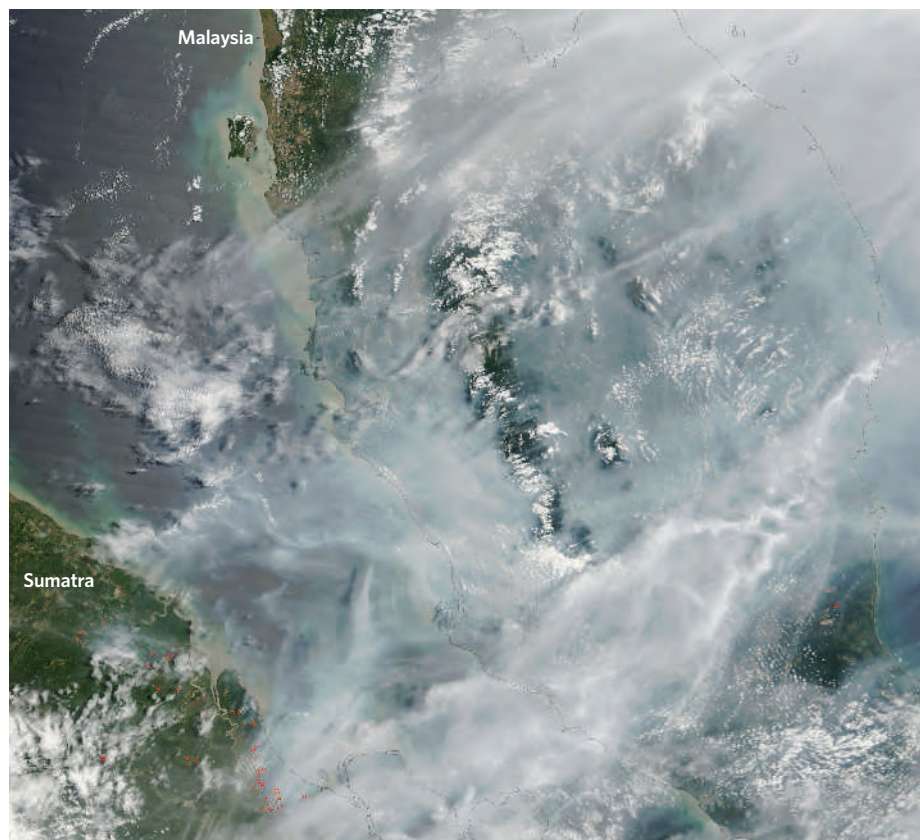
Asia choked by peat smog

Countries in southeast Asia are planning drastic action in a bid to combat peat fires that have smothered the region in a thick haze.

Satellite images taken on 12 August (see picture) show that a soupy smog now covers much of Malaysia and Indonesia. The smog is being generated by smouldering peat bogs on the islands of Borneo and Sumatra, which are set alight each year as farmers clear their lands (see *Nature* 432, 144–146; 2004).

The smoke has been so bad that schools and businesses have been forced to close and there has been a dramatic rise in respiratory ailments, according to reports from the region.

Indonesian authorities announced last week that they would prosecute ten plantation owners for lighting illegal fires on their land. The Malaysian government, meanwhile, is planning to seed rain clouds in the hope that they will help to clear the haze from the air.



J. SCHMALTZ/MODIS RAPID RESPONSE TEAM/NASA GODDARD

Warning system makes the grade during quake

TOKYO

The earthquake that hit northeast Japan last week gave scientists an opportunity to test their early warning system. About 140 schools, government agencies and companies in Sendai, the region's major city, were warned about the 7.2 magnitude quake up to 16 seconds before it arrived.

The quake hit at 11:46 local time on 16 August, about 80 kilometres off the coast of Miyagi state, and caused around 70 injuries. Officials at the Japan Meteorological Agency, which runs the early warning system, say that they are pleased with the accuracy and speed of the warning, although it will still be several years before the system is fully operational. The agency has been testing the system in Sendai since March, and plans to expand it across the country for testing by March next year.

The system takes advantage of a difference in speed between two kinds of seismic waves that make up a tremor. When a quake occurs, hundreds of detectors scattered across the

country pick up the weak primary waves, which move outwards from the epicentre at 6–8 kilometres per second. A computer at the agency's headquarters then calculates how the slower but stronger shear waves, which can cause serious damage, will spread. If the quake is predicted to be more than magnitude 4, it sends out information about the magnitude and estimated arrival time.

"We are now examining how to make the warnings most useful," says Katsuyuki Abe, a seismologist at the University of Tokyo. Apart from warning the public to take cover, the

"There's no other way to protect our lives."

information could be used to stop trains or elevators, or even get surgeons to halt operations.

Many trains and elevators in Japan already have systems to slow them down as soon as a quake hits.

But extra seconds could reduce the amount of damaged track that trains travel over, for example, or give elevators more time to reach the nearest floor.

Japan is one of the world's most earthquake-prone countries. The next 'big' quake, of per-

haps magnitude 8 or more, is expected to hit the nation within the next 30 years. The government's Central Disaster Prevention Council predicts that about 9,200 people would die if such a quake hit off the coast of the Tokai region, west of Tokyo, but officials hope that the early warning system, together with ongoing monitoring of the strain in the tectonic plates, could significantly help to reduce that toll.

The latest success will help build the credibility of the warning system, which has delivered a few false alarms during thunderstorms. But improvements are still needed. At a school in Sendai, the warning signal was meant to be relayed to loudspeakers to tell the children to take shelter, but a computer error meant it didn't work. Fortunately, the children were on their summer holidays at the time.

"It's still at the development stage," admits Kosaku Yamaguchi, a senior researcher at the Real-time Earthquake Information Consortium. "But we are getting to know what the problems are. There's no other way to protect our lives."

Ichiko Fuyuno

ON THE RECORD

“At the end of 2.5 years and \$1.5 billion or more, it is not clear what has been accomplished.”

Some members of a NASA task group question the success of the space shuttle's return-to-flight programme.

“They were cold and wet, had no equipment or weapons and were surrounded by hungry polar bears.”

A Norwegian official describes the plight of three Polish scientists whose boat was wrecked 1,000 km from the North Pole. The trio fended off the bears long enough to be rescued.

Source: *Chicago Tribune*

OVERHYPED

Addiction

There's no such thing as a quick fix for heroin addiction. But some doctors offer fast detoxification — costing up to US\$15,000 — in which the patient is given an addiction-fighting drug while under general anaesthetic. Many have warned against such expensive programmes — and with good reason, says a study in which heroin addicts were given one of three treatments. Those who went through rapid detox fared no better than those getting traditional outpatient drugs, and faced the extra risks of anaesthesia (E. D. Collins *et al.* *J. Am. Med. Assoc.* 294, 903–913; 2005). The way out of addiction remains far more difficult than the way in.

NUMBER CRUNCH

The US National Institute of Allergy and Infectious Diseases surveyed some 3,500 US adults about HIV vaccine research and found:

73% feel it is important to support HIV vaccine research personally.

24% don't know whether vaccines being tested can cause HIV infection — or incorrectly believe that they do.

18% believe an HIV vaccine already exists and is being kept a secret.

Source: M. A. Allen *et al.* *J. Acq. Immun. Def. Syn.* doi:10.1097/01.qai.0000174655.63653.38 (2005).

Alertness drug arouses fears about 'lifestyle' misuse

A drug being studied as a potential treatment for Alzheimer's can also counter the effects of sleep deprivation, a new study suggests. The finding has sparked debate over the use of 'lifestyle' drugs, which people take to make themselves feel smarter or more alert, rather than for a specific medical condition.

Developers of the drug, known as CX717, say it is meant to treat a range of debilitating mental conditions. But bioethicists point out that it could easily follow in the footsteps of other treatments that are being prescribed 'off-label', such as modafinil, a narcolepsy drug that is gaining popularity as a pick-me-up pill.

On 22 August, researchers at Wake Forest University Baptist Medical Center in Winston-Salem, North Carolina, published the results of trials of CX717 in monkeys. Animals that had been deprived of sleep for up to 36 hours — equivalent to 72 hours in humans — experienced cognitive deficits that virtually disappeared after taking the drug (L. J. Porrino *et al.* *PLoS Biol.* doi:10.1371/journal.pbio.0030299). Rested animals that were given CX717 also did better on cognitive tests than control animals. The drug works by boosting the uptake by brain receptors of the neurotransmitter glutamate.

The results in monkeys support findings from an unpublished human trial, funded by the company that makes the drug. In that study, 16 men kept awake overnight did better on a range of memory and attention tests when dosed with CX717; those suffering the most from tiredness received the biggest boost

in performance. “We didn't see any adverse events,” says Julia Boyle, who ran the study at the University of Surrey in Guildford, UK.

The drug's developer, Cortex Pharmaceuticals of Irvine, California, is pushing ahead with trials in patients with Alzheimer's and attention deficit hyperactivity disorder, as well as with people working night shifts. Bioethicists say that the drug may find other 'lifestyle' uses, perhaps by allowing workaholics to work longer days.

In a similar case, modafinil (sold under trade names including Provigil) was licensed in the United States as a treatment for sleepiness caused by narcolepsy in 1998 by Cephalon of Frazer, Pennsylvania, and later for other specific sleep disorders.

If CX717 can keep users awake as well as modafinil, with an added cognitive boost, experts worry that it could be misused by tired office workers or students with essay deadlines. “This could coerce people into staying awake an extra two hours,” says Arthur Caplan, a bioethicist at the University of Pennsylvania in Philadelphia.

The US Defense Advanced Research Projects Agency is already funding studies designed to assess whether CX717 could help soldiers stay more alert.

Roger Stoll, chairman of Cortex Pharmaceuticals, says that his company is focusing only on clinical disorders in which CX717 might help. Still, he admits, “we don't know what everyone will do with it”.

Jim Giles

Ethicists fear that people might misuse a new drug, known as CX717, to work long hours.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

K. KASMAUSKI/CORBIS



**NO SHUTTLE FLIGHTS
UNTIL 2006**

Easy fix for foam problem
eludes NASA.

www.nature.com/news

NASA

R. JONES/PANOS

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Africa urged to create more fish farms

Fisheries experts are this week meeting the leaders of 25 African countries in Abuja, Nigeria, to call for urgent investment in aquaculture across the continent.

Africa's rapidly rising population means that it will need to produce significantly more fish to provide the same amount of food per head as it does now. But wild fish stocks in the continent are already on the verge of destruction.

The warning comes from the World Fish Center, an international research organization based in Malaysia, which has carried out a series of investigations into the state of Africa's fisheries. It finds that Africa is the only region in the world where per capita fish supplies are falling.

In Africa, only Egypt has developed aquaculture fast enough to match population growth, says Daniel Jamu of the World Fish Center. "The rest of Africa still has a long way to go."

The situation is especially serious in sub-Saharan Africa, where the per capita fish supply dropped from 9 kilograms a year in 1973 to 6.6 kilograms a year in 2001. Worldwide, that

number increased from 12 to 16 kilograms a year over the same time period.

Fish is one of the most important sources of nutrition for Africans — it is much cheaper than meat, and almost 30% of the 690 million people in sub-Saharan Africa rely on it as the main part of their diet.

A third of sub-Saharan Africans are already undernourished: vitamin A deficiency contributes to the deaths of around half a million children each year and up to 20,000 women die annually from iron deficiency. Doing nothing about fish production would make these problems much worse, says Patrick Dugan, deputy director-general of the World Fish Center.

To maintain current consumption levels as the population rises, sub-Saharan Africa will need 32% more fish by 2020, says Dugan. Wild fish stocks could collapse if fished any harder — in some areas, such as Lake Malawi, that is already happening. So most of the increase will have to come from fish farms. This means that sub-Saharan Africa would need to produce 3.6 times as many fish from aquaculture by 2020 as it does now — which would require the

construction of thousands of freshwater ponds.

At least eight of the governments attending the Abuja meeting, including Kenya, Uganda and Malawi, are expected to sign a declaration on 25 August calling for help from the international community. Dugan says that US\$30 million every five years should be enough to support an annual 10% increase in fish-farm output. The idea is that manual labour should be sufficient to dig the ponds, and the fish living in them can be fed with organic waste from the farmers' gardens.

The United States, Canada, Britain, Norway, Germany and Japan are already supporting similar efforts in countries such as Niger, Malawi and Uganda, says Richard Mkan-dawire, agricultural adviser for the New Partnership for Africa's Development, a continent-wide development initiative set up by the African Union. But he says he hopes that the conference will lead to increased efforts. "We see this meeting as a turning point for the revival of African fisheries as well as aquaculture development," he says.

Andreas von Bubnoff

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Warm reception: environment ministers were shown the effects of global warming in Greenland.

Denmark breaks the ice over climate-change talks

With the future of the Kyoto Protocol increasingly being questioned, Scandinavian countries last week took the initiative to try to move discussions on climate change forward.

The Kyoto agreement expires in 2012, and nothing has yet been agreed about its successor. Talks will continue in November in Canada, where the countries participating in the Kyoto agreement will start negotiating new targets for cutting greenhouse-gas emissions.

As an early step, the Danish environment ministry last week organized the informal 'Greenland dialogue' in the remote town of Ilulissat. Environment ministers from 22 countries toured vanishing glaciers and ice fjords to see Arctic climate change for themselves.

No formal decisions were taken in Greenland, but organizers say that the meeting was a step forward in breaking the impasse between countries. "We have now agreed to look ahead and stop blaming each other for not solving climate problems," says Connie Hedegaard, the Danish environment minister.

Archaeologist shot dead by robbers in Brazil

Archaeologist James Petersen was shot and killed by robbers in Iranduba, Brazil, near where he was doing fieldwork, on 13 August.

Petersen, who was chair of the anthropology department at the University of Vermont, Burlington, was known for his theory that the rich black soil found in some parts of the Amazon was deliberately created by humans for farming. This idea stood in opposition to the reigning 'counterfeit paradise' theory, which held that amazonian soils could not support dense or advanced populations. Petersen's theory gradually won favour as he found pottery shards and

other evidence of civilization in the soils.

He was killed in a restaurant where the robbery took place; three people have been arrested. News reports say the suspects are in their teens and were high on cocaine and alcohol at the time.

Russian rocket set to give Europe's science a lift

Another agonizing delay to the space shuttle's flight schedule has led the European Space Agency (ESA) to turn to Russia for help with its research programme.

In a deal negotiated before the next shuttle flight was delayed until March 2006, ESA has reshuffled its plans for 12 of the 30 experiments that were due to be carried out by German astronaut Thomas Reiter. Reiter will fly on the next shuttle for a six-month stay at the International Space Station.

The experiments will be done instead by a Russian cosmonaut sent up on a Soyuz rocket in the first week of October. ESA will pay an undisclosed fee for the cosmonaut's time and will share the data from the mainly physiological experiments with the Institute for Biomedical Problems in Moscow.

ESA hopes the remaining experiments will simply be delayed by six months. The Russian rocket is not capable of lifting into orbit the stranded 13-tonne European science module Columbus, which must wait its turn for the shuttle.

Travel limits cause agency to halt aid to Myanmar

Public-health grants worth around US\$36 million are being withdrawn from Myanmar after the nation's authorities

imposed travel restrictions on aid officials.

The Global Fund to Fight AIDS, Tuberculosis and Malaria pulled out on 18 August after Myanmar's government restricted access to key areas. The southeast Asian nation, which is ruled by a strict military regime, was originally targeted by the fund because of its location between three countries with serious HIV problems: India, China and Thailand. Myanmar also has one of the highest rates of tuberculosis in the world, with almost 100,000 new cases detected every year.

The fund says that its programmes in Myanmar will be phased out by the end of the year and that a "large part" of the \$12 million that has already been granted will be recovered.

Supercomputer grid nets cash boost for expansion

TeraGrid, the world's largest distributed computing network for scientific research, has been awarded US\$150 million by the US National Science Foundation to keep it growing over the next five years.

The network, which was completed in September 2004, connects 16 supercomputers across the United States. It can run more than 50 trillion calculations per second and has been used in many different projects that require a lot of data-crunching. Scientists have used TeraGrid to simulate protein folding, model earthquakes, and forecast the weather, among other studies.

The cash will allow more scientists and engineers to access TeraGrid from their desktop computers through user-friendly software, says Charlie Catlett, the project's director.

Student uncovers Einstein's long-lost manuscript

A Dutch graduate student has unearthed a handwritten version of a paper by Einstein.

The student stumbled across the 16-page German manuscript — *Quantum Theory of the Monatomic Ideal Gas* — while searching the Paul Ehrenfest archives at the University of Leiden's Lorentz Institute for Theoretical Physics. Einstein frequently stayed at Ehrenfest's house when he was invited to lecture at Leiden.

Einstein's notes on the manuscript (right), which predicts a state of matter now known as a Bose-Einstein condensate, suggest that he used it to correct proofs from the Prussian Academy of Sciences for a paper published in 1925.

Einstein later developed his idea — that atoms could form a superfluid phase at temperatures near absolute zero — with the Indian physicist Satyendra Bose. The first such condensate, in dilute gases of alkali atoms, was produced by Eric Cornell and Carl Wieman in 1995 at the University of Colorado, Boulder.

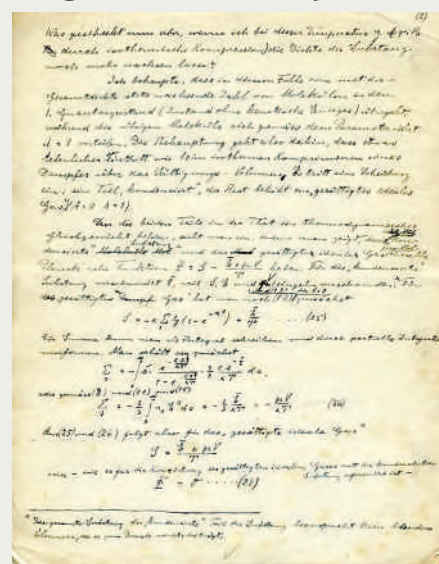


IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Knitting pattern: the same process seals the eyes of mice before birth and heals wounds in embryos.

The healing touch

Wound an embryo and it heals perfectly, with no scars. Can we teach adult wounds the same trick, asks **Meredith Wadman**.

It was alligators that first drew Mark Ferguson to study human scarring, but not for the reasons you might expect. In the late 1970s, he had become fascinated by cleft palate. Alligators were the obvious research subject: their embryos have palates, and develop in easily accessible eggs. But when Ferguson performed surgery on alligator embryos to mimic cleft palates, the creatures hatched with completely normal, unscarred mouths. "As a surgical model of cleft palate it was perfectly useless," says Ferguson. "As an observation of scar-free healing, it was of great scientific and clinical interest."

Ferguson, who now researches wound healing at the University of Manchester, UK, was not the first to stumble across this phenomenon. As early as 1960 there were anecdotal reports that wounds made early in gestation in embryos of many species, including humans, heal rapidly and perfectly. Over the past 20

years, researchers have been probing the mysteries of this process in the hope of improving adult wound healing and perhaps even making scarless healing a reality. It's no trivial point: delayed wound healing in the US elderly, for instance, is estimated to cost more than \$9 billion each year.

Growing realization

Such work has wider implications too. There are striking similarities between the mechanisms embryos use to heal wounds and those they use to knit their body parts together during normal development. Research is throwing light on these basic processes and what might go wrong with them to cause birth defects such as cleft palate.

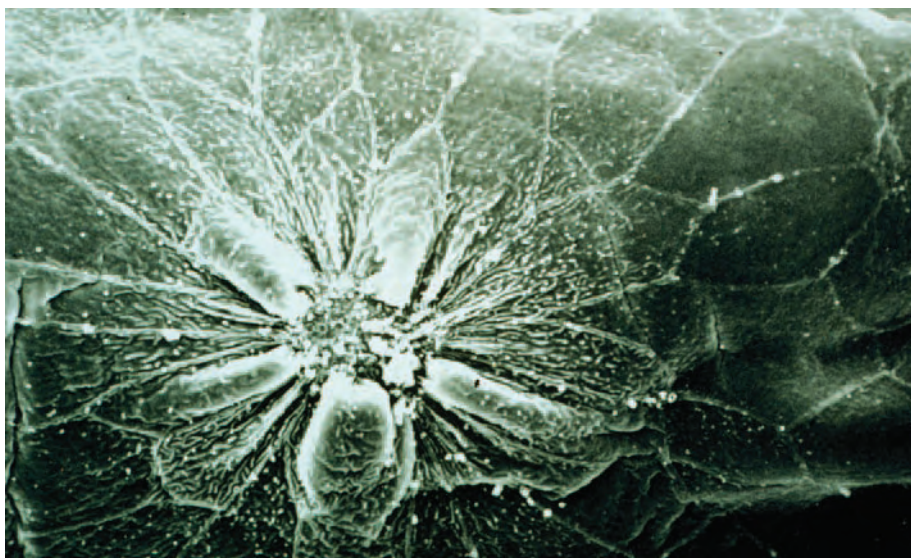
When adult skin is wounded, a blood clot quickly forms to stem the bleeding, then cells underneath move in for the repair job. Some of these cells, called fibroblasts, churn out a

matrix of support proteins. Skin cells at the edge of the wound drag themselves over this to close the wound. As immune cells rush to the scene, the underlying connective tissue contracts, leading to scarring.

Embryos do things very differently, says Paul Martin, a cell and developmental biologist at the University of Bristol, UK. A fruitfly embryo closing a wound rapidly assembles a cable of the protein actin in the 'skin' cells at the wound's edges. Contraction of this cable seems to pull the edges together much as a drawstring pulls a bag shut¹. Once the edges meet, the cells reach out with thin extensions called filopodia, which interlock to seal the wound. Flies unable to make these cell extensions cannot close wounds completely².

Intriguingly, this tissue repair uses the same machinery used by embryos to close holes that normally arise as they build their bodies, says Martin. This makes sense, he says: embryos are in the business of building tissue, and wound healing is essentially a process of rebuilding what was damaged. To close a wound, he says, "all they do is reactivate the machinery they're using somewhere else in the body anyway".

For example, a developing fruitfly embryo needs to zip up its outer skin as this grows over its back. This process, dubbed dorsal closure (see graphic), uses the same actin cable and filopodia mechanism. What's more, this mech-



On the mend: a three-day-old zebrafish embryo draws a wound shut on its outer layer.

anism is remarkably common across the animal kingdom, whether it's in an embryonic mouse knitting its eyelids shut or a worm fusing its outer surface. It is also thought that filopodia are critical for a human embryo to fuse its palate or the tube that forms its brain and spinal cord.

Wrong turning

When these events go wrong in humans, notes Martin, the consequences range from the serious, such as cleft palate, to the devastating, as in anencephaly, when a baby is born missing a large part of its brain and skull. So it is hoped that understanding the process better will mark a beginning in the long journey towards preventing this kind of problem.

The evolutionary conservation of the healing mechanism means that researchers can study its basic biology easily in lab workhorses such as flies and mice. One focus is a signalling pathway between cells known as the JNK cascade, which plays a central role in skin-cell migration both after wounding and during normal development³. The genes switched on by this cascade are activated during dorsal closure in flies and within minutes in wounded rat embryos, in response, it is thought, to mechanical stress on cells.

Biologists such as Martin would dearly love to know more about how to activate this signalling, to kick-start healing in chronic wounds such as diabetic leg ulcers. Recent findings that conserved genes also control skin repair once wounds are closed in flies and mammals offers another target for therapy^{4,5}. For the moment, such treatments are still a distant dream. But there is one facet of embryonic wound healing that could have clinical relevance sooner: inflammation.

Wounds in adults are red and inflamed, bristling with immune cells summoned by molecular signals from platelets that leak from damaged blood vessels. These signals are also thought to trigger the skin-cell migration and connective-tissue contraction that ultimately

lead to scarring. The inflammatory response is a robust one, which makes sense in the face of invading bacteria. Scarring could be the price to pay for protecting against infection.

Embryos cannot mount such a response until late in gestation, and this could be part of the reason they don't scar. Martin and his colleagues recently studied mice genetically incapable of raising an immune response because they lack key immune cells called macrophages and neutrophils⁶. When these mice are wounded as newborns, they not only seem not to scar, they also heal faster.

Gillian Ashcroft, a consultant in tissue repair at the University of Manchester, UK, further bolstered the case when she showed that mice lacking a protein called Smad3, which is involved in inflammation, healed their skin wounds faster than normal mice⁷. Immunologist Luisa DiPietro and her colleagues at Loyola University Medical Center in Maywood, Illinois, have made similar findings in mice lacking neutrophils⁸.

DiPietro's team has also looked at the

mucous membranes in the mouth. Although this environment is a hotbed of bacteria, wounds in the mouth of mice show less inflammation and heal faster than those in the skin⁹. "The emerging evidence all leads to this idea that in healing wounds, less inflammation might be better," says DiPietro.

To hear Ferguson tell it, the drive to better healing is already under way. In 2000, he co-founded Renovo, a Manchester-based biotechnology company. The firm is well into clinical trials of several drugs aimed at reducing post-operative scarring.

Mimicking embryos

Renovo's drugs stem from Ferguson's work on different forms of a chemical signal, a protein called transforming growth factor β (TGF β). This plays a key role in scar formation, summoning and activating inflammatory cells, as well as spurring fibroblasts to make matrix. In work on mice and sheep embryos, Ferguson found that embryonic wounds have very high levels of one form of the protein, called TGF β -3, but very low levels of TGF β -1 and TGF β -2. In adult wounds, the profiles of these forms are reversed^{10,11}. Ferguson's team discovered that by mimicking the embryonic profiles in adult animal wounds through adding more TGF β -3 or suppressing TGF β -1 and TGF β -2, they could reduce scarring¹².

Today, Renovo's flagship product — a human genetically engineered TGF β -3 dubbed Juvista, has passed initial safety and efficacy trials in skin wounds. Ferguson is now preparing to publish the trial data, and if all goes well, he predicts that Juvista could be on the market in 2009.

Martin, too, is optimistic. He foresees a day when designer drugs modulating the inflammatory response will improve myriad diseases in which inflammation goes awry, from Crohn's to heart disease.

Getting there, he notes, will take a far better understanding of what he calls "this can of worms we know nothing about". He adds: "Because this is such a complex process, we've got to go back to the embryo, back to a genetically tractable organism like the fly, to start to understand it and to watch it."

Meredith Wadman is a freelance writer based in Washington DC.

Drawing on embryos

The way fruitfly embryos 'zip up' their outer skin could give new insights into wound healing.

Head Tail

A cable of protein called actin (orange) forms a ring in the cells at the edge of the hole.

Thin protrusions (blue) then help to knit the skin together.

1. Martin, P. & Lewis, J. *Nature* **360**, 179–183 (1992).
2. Wood, W. *et al.* *Nature Cell Biol.* **4**, 907–912 (2002).
3. Martin, P. & Parkhurst, S.M., *Development* **131**, 3021–3034 (2004).
4. Mace, K. A. *et al.* *Science* **308**, 381–385 (2005).
5. Ting, S. B. *et al.* *Science* **308**, 411–413 (2005).
6. Martin, P. *et al.* *Curr. Biol.* **13**, 1122–1128 (2003).
7. Ashcroft, G. S. *et al.* *Nature Cell Biol.* **1**, 260–266 (1999).
8. Dovi, J. V., He, L.-K. & DiPietro, L. A. *J. Leuk. Biol.* **73**, 448–455 (2003).
9. Szpadarska, A. M., Zuckerman, J. D. & DiPietro, L. A. *J. Dent. Res.* **82**, 621–626 (2003).
10. O'Kane, S. & Ferguson, M. W. J. *Int. J. Biochem. Cell Biol.* **29**, 63–78 (1997).
11. Whitby, D. J. & Ferguson, M. W. J. *Dev. Biol.* **147**, 207–215 (1991).
12. Shah, M., Foreman, D. M. & Ferguson, M. W. J. *Lancet* **339**, 213–214 (1992).

In their element

At first it was just an unusual, geeky hobby. But by combining their twin passions of chemistry and history, Jim and Jenny Marshall are now running an acclaimed project in science education. **Alexandra Witze** reports.



Not every honeymoon turns into a scholarly project. But that's what happened when Jenny Marshall told her husband, chemist Jim Marshall, that she'd like to spend the summer after their wedding visiting European sites of chemical history.

Six years later, that unusual honeymoon trip has spawned an unparalleled historical investigation. From their home in Texas, the Marshalls travel across Europe every summer to the places where various chemical elements were discovered. They have amassed hundreds of photographs, original documents and many colourful anecdotes — from boating along the coast of Norway in search of thorium, to navigating a muddy Transylvanian track to reach the original tellurium mine.

Few, if any, have stood in quite so many spots where one of the elements was first discovered. "It makes the hairs rise on the back of your neck as you realize this is where history was made,"

says Jim. And as public interest in chemistry flags, the Marshalls see their work as a way to stimulate excitement about a field so rich in history. "Chemical education can be too junior-high-schoolish, with lots of loud bangs and prettily coloured gases," says Jim, "and history can be awfully stuffy and plain and boring."

The human factor

The Marshalls' project is anything but, experts agree. "It gives this wonderful human reality to how chemistry became what it is today," says Alan Rocke, a historian of science at Case Western Reserve University in Cleveland, Ohio. "Scientists are sometimes too ready to ignore their history and the wider culture of their field, and I think that's a loss." Jim Marshall is determined to bridge that culture gap, lecturing regularly on the American Chemical Society's speaker circuit. His web page has become a staple resource for students working on projects about a particular element. And with his wife,

he is compiling a DVD called *Rediscovery of the Elements*, with photos and narratives about their quest, scheduled for release in 2007.

Neither of the Marshalls is a professional historian. Jim teaches chemistry at the University of North Texas in Denton; Jenny is retired from teaching computer technology in local middle schools. Jim's research focus is on organic chemistry and materials science. During the 1980s he worked in industry, including a stint at Motorola that proved key to his leisure-time interests.

At Motorola, Jim began collecting samples of as many elements as he could find. His industry contacts brought him some hard-to-find materials, including promethium from a nuclear reactor at the Oak Ridge National Laboratory in Tennessee. Marshall eventually gathered examples of every element from hydrogen to uranium — the latter represented by armour-penetrating depleted uranium from the tip of a missile stockpiled for the first Gulf War.



Metal mix-up: the Marshalls have shown how a chemist's carelessness meant vanadium was discovered twice

The collection fills a wall of bookcases in the living room of the Marshalls' small town-house. Glass bubbles filled with gas share the shelves with combustible phosphorus, submerged in oil. "I'm afraid to dust because I don't know what will blow up," jokes Jenny.

Historic honeymoon

Given Jim's collection of elements, it seemed natural for Jenny to suggest visiting the sites at which some of them were discovered when the Marshalls made plans for their 1999 honeymoon. That first trip set the pattern: at each site, the Marshalls go forth with historical maps to find the place of interest, cameras to shoot it, and Global Positioning System receivers to fix the location. They then try to secure a new sample, often in the form of a mineral containing the element in question. These 'rediscovery' samples now enrich the original collection.

In many cases, the rediscovery work brought to light stories that were known locally but not widely among historians of science. In Norway, the Marshalls visited mineralogist Alf Olav Larsen for a boat tour along the Langesund Fjord, south of Oslo. They motored along the craggy shoreline, just as Hans Esmark, a local pastor, had done in 1829 while hunting ducks. And they visited a rock outcrop from which Esmark may have chipped a shiny black mineral for the first time. The Swedish chemist Jöns Jakob Berzelius later named that mineral thorite after the Norse god of thunder, and isolated the element thorium from it.

In Romania, the Marshalls travelled to the remote Fata Baii mine in search of tellurium. Franz-Joseph Müller von Reichenstein had identified the element in 1783, after puzzling over what he called a "*metallum problematum*" for several years. Working from descriptions written in the early nineteenth century by a British mineralogist, the Marshalls identified the original mine from which Müller von Reichenstein had obtained his materials. Local

guides led them up a steep, muddy road into the Transylvanian forests, where they finally came across the entrance to the ancient mine, framed by rotting timbers.

Not all the journeys were so arduous. For gallium, the Marshalls tracked down the former home of Paul-Émile Lecoq de Boisbaudran, whose family owned a wine business in the Cognac region of southwestern France. In a well equipped home laboratory, Boisbaudran used spectroscopy to identify an unknown element in an ore mined from the Pyrenees. It melted in his hand — a discovery that delighted Russian chemist Dmitri Mendeleev, who had predicted its properties with his recently developed periodic table of the elements.

In some places, the Marshalls had to dig strenuously to hit pay dirt. In Germany, they eventually found that the laboratory where indium was first isolated is now a bathroom at the Freiberg Academy of Mining and Technology. In Paris, the mining school where chromium and beryllium were discovered had become a children's clothing shop. But even in such places, says Jim Marshall, "you can just smell the history".

In other areas, the Marshalls ran into long-standing disputes over who should be properly credited with an element's discovery. Does the person who first isolated it deserve recognition? Or the one who realized its importance? Thallium, for instance, was first seen in spectroscopic experiments by William Crookes in London. But French physicist Claude-August Lamy cast an ingot of it the next year, grabbing public acclaim for the discovery until Crookes protested loudly and won joint credit.

"Scientists are sometimes too ready to ignore their history and wider culture, and that's a loss." — Alan Rocke

The Marshalls have uncovered new details of a similar misunderstanding over the discovery of vanadium. Historians of science know the tale of Andrés Manuel del Río, a mineralogist in Mexico City who isolated vanadium from Mexican ore. He gave some samples of the new element to explorer Alexander von Humboldt, who took it to Europe and passed it to a French chemist for analysis. Upon testing, he became convinced that the Mexican element was chromium, which like the unknown ore also produced a range of brilliantly coloured salts. And on hearing the findings, del Río retracted his claim to have discovered a new element.

Jim Marshall recreated the French tests in his kitchen sink and found that the chemist had ignored an obvious inconsistency in the colours of certain precipitates. Had he paid closer attention to the colours, Marshall argues, he would have realized that del Río's substance was in fact an unknown element. Instead, it took another three decades before vanadium was 'discovered' by a Swedish chemist. Fortunately, however, del Río did eventually get retrospective credit for his find.

In another case, evidence amassed by the Marshalls may mean history needs rewriting. The discovery of radon is typically attributed to Friedrich Dorn, who in 1900 studied a gas built up inside capsules containing compounds of radium. But the original paper had been incorrectly cited for decades, and its contents drifted into obscurity. So the Marshalls went to Halle, Germany, to find the journal article.

Radon rewritten

Working from the original paper, they decided that Dorn hadn't correctly understood the "emanation" that built up. In the previous year, New Zealander Ernest Rutherford had made similar observations of a different isotope of radon emitted by a sample of thorium, and later to be dubbed thoron. Because he characterized the emanation fully, and placed it correctly in the periodic table, "Rutherford should be given credit for the discovery of radon", the Marshalls wrote in a 2003 paper (J. L. Marshall and V. R. Marshall *Bull. Hist. Chem.* **28**, 76–83; 2003). "Personally, I think they've made an absolutely convincing argument," says Rocke. But he warns that it may take some time for the textbook version of events to incorporate the lessons of the Marshalls' research.

The Marshalls are pleased to have made an original contribution to the history of chemistry, but say that their favourite element remains one whose story is well known: radium, painstakingly isolated by Pierre and Marie Curie from 10 tonnes of uranium ore. Jenny speaks in hushed tones about the time she visited their laboratory, and was allowed to try on one of Marie's smocks. "The story's just so romantic," she says. There speaks a woman whose idea of a dream honeymoon is a voyage of chemical rediscovery. ■

Alexandra Witze is a senior news and features editor for *Nature*, based in Washington DC.

Seeking the solution

Is there any fundamental reason to be fixated on water as the universal elixir of life?

Philip Ball investigates.

Where there's water, there's life. That, at least, is what our experience on Earth has taught us, and when it comes to searching for life on other worlds, NASA seems determined to follow the water. But is it right to see water as the sole medium for extraterrestrial life?

Some think not. "Water is a terrible solvent for life," says chemist Steven Benner of the University of Florida in Gainesville. Benner is one of a number of biochemists, planetary scientists and philosophers who are trying to find out whether water is in some sense 'fine-tuned' for life.

It is a pertinent question. Scientists today are searching for extraterrestrial life wherever they can, from the subsoil of Mars to planets orbiting other stars. So far they have found nothing. But are they looking in the right places? If Benner is right, perhaps they have become too fixated on chasing water.

Benner led the case for the prosecution earlier this year at a meeting in Varenna, Italy. There, researchers faced up to what many of them consider to be their biggest challenge. How can we even begin to discuss the relationship between water and life when we have only one example — life on Earth. And more to the point, can the problem be dealt with in a rigorous, scientific manner?

For Benner the answer to the latter question is yes. He argues that it is possible to investigate experimentally whether water is essential for life. He hopes to prove that a type of biochemistry can occur without water. "We are working to create alternative darwinian systems based on fundamentally different chemistries," he says. "We are using different solvent systems as a way to get a precursor for life on Earth."

The notion of redesigning life's chemistry has become central to the emerging discipline of synthetic biology, which has among its long-term objectives the aim of creating entire cells from scratch, perhaps with a different chemical basis from that of existing organisms.

Benner points out that water is generally not a good solvent for doing organic chemistry — which is, in the end, what life is all about. For one thing, water is rather reactive, tending to split apart the bonds that link the building blocks of biomolecules together. It readily breaks peptide bonds, for example, as well as many of the bonds in nucleic acids, such as RNA. "The structure of RNA screams 'I did not arise in water!'" Benner asserts. He says that in about four out of five cases, synthetic organic chemists will avoid using water as a solvent.

Creative force

But of course organic chemists aren't usually trying to create life. Water has many properties that seem indispensable for the functioning of proteins and cells. It is an excellent solvent for ions, for example — crucial for nerve signalling, enzymatic processes, biomineralization and the behaviour of DNA. It is also a master of weak intermolecular interactions such as hydrogen bonds and hydrophobic forces. The latter play a central role in protein folding and protein-protein interactions, whereas the former often act as bridges between protein binding sites and their substrates. And water's ability to absorb and lose heat without undergoing a large temperature change provides thermal cushioning, shielding cells and organisms from wild temperature swings.

No other known liquid combines all of these properties. But does a life-supporting solvent need them all? Are any of water's unique prop-

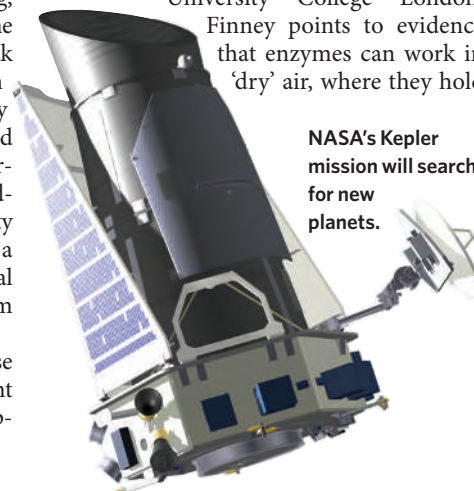
erties essential, and are any of its essential properties unique?

To assess whether water is somehow 'special' as a biological solvent, we need to understand the basic requirements for life¹. Proteins and nucleic acids rely on weak intermolecular interactions to organize and pass chemical information around — to transfer genetic instructions, for instance. It could be argued that general properties such as this will be needed for a 'chemistry of life', even when the building blocks are not proteins or nucleic acids.

But these familiar building blocks may themselves not need water to function. "I think it is perfectly possible that at least elements of relevant biochemistry can be persuaded to work in a completely non-aqueous environment," says physicist John Finney at University College London.

Finney points to evidence that enzymes can work in 'dry' air, where they hold

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS



NASA's Kepler mission will search for new planets.

F. SAUZE/SPL

NASA

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS



The clouds of Jupiter contain liquid ammonia which some say could support water-free life.

on to only the barest coatings of water molecules, and even in non-aqueous solvents².

Most natural enzymes cannot fold into their compact, active forms without being immersed in water. But even that limitation might not be fundamental. A team led by Annelise Barron and Ishwar Radhakrishnan at Northwestern University in Evanston, Illinois, has recently found that molecules called peptoids, which are structurally very similar to peptides, can fold into compact forms in pure methanol.

Another way to explore the relationship between water and life is to modify water's molecular structure or properties until the liquid state itself begins to change. "Knowing which properties of water are particularly sensitive to its structure can help to show how fine-tuned for life the liquid properties are," says Ruth Lynden-Bell, a theoretical chemist at Queen's University Belfast, UK, who co-chaired the Varenna meeting.

Lynden-Bell and her co-workers have used computer simulations to model changes in water's properties. They found that if the bond angle in H₂O was 90°, rather than 104.5° as in the real molecule, or if the hydrogen bonds were about 15% weaker, the three-dimensional network of hydrogen bonds — crucial to the liquid's unique properties — would be severely disrupted or fall to pieces^{3,4}.

Asking such 'what if' questions might seem strange to biologists and chemists, but it is far more common in cosmology or physics. For

cosmologists, the physical Universe seems to be precariously fine-tuned to make life possible⁵. For example, the fine-structure constant, which determines the strength of electromagnetic interactions, is not fixed by any known fundamental theory; and yet if it was ten times larger, stable atoms could not exist.

Unlike physics, of course, biochemistry adapts to its environment, which is why the participants in Varenna generally agreed that life on Earth is adapted to water rather than the other way round. "Life on Earth itself is fine-tuned to water — a consequence of it evolving in close association with the medium," says Finney. "To put it the other way is perhaps to put the cart before the horse." He adds that "the fine-tuning argument with respect to water is a far more complex problem than that in astrophysics. Without knowing what aspects of water are important, I suspect we are doing little more than speculating."

Out of this world

Benner wants to use experiments to move beyond these abstract discussions. He sees several alternatives to water⁶. Ammonia, which is liquid between -78 °C and -33 °C at 1 atmosphere pressure, will dissolve many organic compounds and can form hydrogen bonds. It is also relatively common in the cosmos: there is liquid ammonia in the clouds of Jupiter, for example. Benner considers water-ammonia mixtures, which may exist in some cold extraterrestrial environments such as beneath the surface of Saturn's moon Titan, as another promising candidate. Then there is

formamide, which is liquid over a wide range of temperatures and pressures, dissolves salts and has hydrophobic-like effects. Formamide might be present below the surface on Mars.

But Benner does not rule out more exotic possibilities, such as liquid nitrogen or supercritical liquid hydrogen on gas-giant planets such as Saturn, Uranus and Neptune. Or perhaps hydrocarbons such as liquid methane on Titan. "Organic reactivity in hydrocarbon solvents is no less versatile than in water," he says.

Does widening the search for extraterrestrial life to places that don't have water make sense? The Cassini-Huygens space probe showed in January that Titan may have riverbeds (if not actual rivers) carved from liquid hydrocarbons. And in March the detection of reflected light from two Jupiter-like extrasolar planets by the Spitzer Space Telescope⁷ marked the first step towards analysing their chemistry remotely.

But NASA's quest for habitable planets remains focused on water. In 2008 it intends to launch the Kepler Photometer, which will search for Earth-like planets by looking for evidence of their transit across the faces of parent stars. At a much later date, the agency hopes to launch two space-based telescopes that will form the Terrestrial Planet Finder (TPF) mission. These will detect and analyse reflected light from other planets. The gaze of both Kepler and the TPF will be firmly fixed on the 'habitable zone' of stars, where liquid water could exist, potentially overlooking worlds that are habitable by non-aqueous life forms.

Benner is not waiting around for these space missions to find extraterrestrial life in places with or without water. He is convinced that the time is ripe to explore more exotic life forms in the laboratory. But that, he says, requires a different mindset from the one that currently guides chemical research and funding. Benner is participating in a US National Academies panel funded by NASA that is looking at possible alternative chemistries for life, and which he hopes will identify research directions that funding agencies can pursue. He believes that researchers should aim high — to create life forms that do not reproduce the chemistry that is found on Earth. In other words, if we can't easily get to other worlds, we should build them here. ■

Philip Ball is a consultant editor for *Nature*.

"Life on Earth is fine-tuned to water — a consequence of it evolving in close association with the medium."

— John Finney

1. Benner, S. A. & Ellington, A. D. *CRC Crit. Rev. Biochem.* **23**, 369–426 (1988).
2. Lee, M.-Y. & Dordick, J. S. *Curr. Opin. Biotechnol.* **13**, 376–384 (2002).
3. Bergman, D. L. & Lynden-Bell, R. M. *Mol. Phys.* **99**, 1011–1021 (2001).
4. Lynden-Bell, R. M. & Debenedetti, P. G. *J. Phys. Chem. B* **109**, 6527–6534 (2005).
5. Hogan, C. J. *Rev. Mod. Phys.* **72**, 1149–1161 (2000).
6. Benner, S. A., Ricardo, A. & Carrigan, M. A. *Curr. Opin. Chem. Biol.* **8**, 672–689 (2004).
7. Deming, D., Seager, S., Richardson, L. J. & Harrington, J. *Nature* **434**, 740–743 (2005).

BUSINESS

Bridging the gap in the German economy

Germany's economic engine has seemed to be stalling of late — partly due, some say, to a breakdown in relations between science and industry.

One man in an excellent position to bring the two together is Günter Stock, head of research at Berlin-based pharmaceutical firm Schering. A physiologist with extensive contacts in academia and industry, Stock takes over next January as head of the Berlin-Brandenburg Academy of Sciences.

Stock is already pondering how best to use his industrial experience to spruce up the academy. In the past few decades, he says, German society has lost its previous consensus that wealth and economic growth are closely linked to progress in science and technology. "Unfortunately, science in this country is often considered a troublemaker rather than a problem solver," he says.

Current angst over Germany's industrial strength contrasts sharply with the optimism of 1982, when Stock joined Schering's cardiovascular research unit. He traces the change, in part, to anti-science sentiment in the wake of the 1980s environmental movement and the Chernobyl accident. He plans to use his new role to forcefully remind the powers-that-be in politics, industry and finance that driving Germany's economy demands the re-engagement of science at all levels.

Expertise where it counts

In a country that lacks a national scientific academy, Stock thinks the Berlin academy can become a flagship institution for channelling expert scientific knowledge to focal areas of society. He aims to bring together expertise from different sources in ways that will be useful to society and government alike. In the past, he says, an array of diffuse opinions has too often confused the public, rather than helping to enlighten it.

Germany's initially cautious response to biotechnology is seen by Stock as an example of where the country has been going wrong. Even when the government launched the BioRegio competition in 1995, which opened door for many biotech entrepreneurs, it ended up with an "over-engineered" regulatory system whose "safety bureaucracy" has continued to stifle the sector's development, he says.

He hopes these mistakes won't be repeated

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Joined-up thinking: Berlin academy head Günter Stock is urging science to re-engage with industry.

in emerging areas, such as nanotechnology. "What's needed is to carefully govern things in the process of doing them, rather than setting up rules and regulations in advance," he says.

Stock thinks that Germany's basic scientific landscape is in good shape, but that there's a divide between publicly funded researchers at the universities and Max Planck Institutes, and their colleagues in industrial laboratories. To bridge this, he hopes to make it easier for scientists to move between the two worlds. Federal and regional government could also help by setting up programmes to foster collaborations between industry and academia. He would like to see the government do more to support small research-based businesses, given the scarcity of private venture capital for such projects in Germany. "Seed capital doesn't like to flow in risky early phase research," he says.

Stock says that after next month's election, he will call on the new government to create a legal framework allowing academia and industry to grasp opportunities early on. He foresees big knock-on benefits: investment in science and technology will also help combat Germany's foremost social problem — unemployment.

His first task, he says, will be to update the academy's interests and long-term research projects. The most burning issues, he thinks, are the problems of an ageing population, molecular medicine and health, nanotechnology, and — in the social sciences — Islamism and conflict research. "We have a lot to offer," he says. ■

Quirin Schiermeier

DPA PICTURE-ALLIANCE/ZB

IN BRIEF

WEIGHT PROBLEMS The US Food and Drug Administration (FDA) has told a watchdog group that it will not ban Abbott Laboratories' Meridia, a weight-loss drug with \$300 million in global sales last year. Back in 2002, US consumer group Public Citizen petitioned the regulator to pull the drug from the market, arguing that its cardiovascular risks outweighed its benefits. But in a 9 August letter to the group, the FDA said that it couldn't attribute the heart attacks and strokes that have been reported in some patients to use of the drug, as such events are so common in obese patients anyway.

SELLING CHIPS Measurement company Agilent, based in Palo Alto, California, has said it will shed several businesses and concentrate on its core activity of making scientific instruments. The company, which was spun-off from Hewlett-Packard in 1999, lost billions of dollars when the high-tech bubble burst and it has since shed some 12,000 staff. Now it is selling its microchip business and retreating from a joint venture with Royal Philips Electronics that makes light-emitting diodes. Bill Sullivan, its chief executive, predicted that the global market for scientific instruments — now worth about \$40 billion, of which Agilent has a large share — will grow by about 8% each year.

BAD FOR BOSTON

Massachusetts-based Boston Scientific, which makes medical devices, says it will stop making a drug-release device after receiving a second warning letter from the US Food and Drug Administration saying that the manufacturing procedure isn't good enough. The Vaxcel device is implanted in the chest, where it dispenses medicine in precise amounts. A previous letter to the same effect was sent in May. Also, studies published in *The Journal of the American Medical Association* and *The New England Journal of Medicine* have reported that the company's leading product, its drug-eluting Taxus stent, is not as effective in preventing the relogging of coronary arteries as its arch-rival Cypher, made by Florida-based company Cordis, owned by Johnson & Johnson.

Dual role for Pluto in the great planetary debate

SIR — The discovery of 2003 UB313, an object larger and farther away than Pluto, has once again stimulated the debate on how we define a planet (*Nature* **436**, 616; 2005). The official status of 2003 UB313 will be decided by the International Astronomical Union (IAU). Meanwhile, in the manuscript for a forthcoming book, I have just changed a section heading from “Eight planets?” to “Ten planets!”. But how should we decide how many planets there are?

I believe that Pluto should remain the ninth planet, as it has been for the past 75 years, and we should enjoy teaching about how this planet differs from the others. Being different is not reason enough to exclude Pluto from the list of planets: in recent years Uranus and Neptune have been found to differ greatly from Jupiter and Saturn, which themselves differ greatly from the terrestrial planets.

I believe that 2003 UB313 should be called the tenth planet, because it is both larger than Pluto and at an appreciably different distance, although a practical problem the IAU will then have to face is where to draw the line at the lower end of sizes. For example, should Sedna or 2003 EL61 — each roughly three-quarters the size of Pluto — also be named planets? I propose that the size of Pluto should be considered the lower limit, for historical reasons.

It is important to remember that Pluto and 2003 UB313 are also ‘Trans-Neptunian objects’ — bodies orbiting the Sun at a greater distance than Neptune — and that these differ from the eight large planets, especially in their origin as small asteroidal aggregates. In order to preserve this distinction, Pluto and 2003 UB313 should also be given asteroid identifications.

Tom Gehrels

Department of Planetary Sciences,
University of Arizona, Tucson,
Arizona 85721, USA

Extra controls will waste yet more research time

SIR — Your Editorial “Rules of engagement” (*Nature* **436**, 2; 2005) discusses the need for biologists to adhere to new codes of conduct. But is further control of biological research really necessary?

In recent years, some urgently needed controls have been adopted. In Brazil, all projects that will involve human or animal experimentation must be submitted to an ethical or animal-use committee for consideration, and any research using genetically modified organisms must be

approved in advance by a federal committee.

In addition, most Brazilian research is preceded by an application for funding, which ensures that projects are refereed by two or more peers before approval. Any resulting manuscripts submitted to accredited journals are again reviewed, with referees in many cases being specifically asked to judge all possible outcomes or uses of the research. In my institute, professional practice in all laboratories is also evaluated by a panel of internal and external investigators every four years.

There is a danger that implementing further controls would simply add to the number of time-consuming obstacles that already make competitive research all but impossible.

Vera Bongertz

Lab AIDS & Imunologia Molecular/IOC/
FIOCRUZ, Avenida Brasil 4365,
21045-900 Rio de Janeiro, RJ, Brazil

Complex research on sea lions is worth the expense

SIR — Your News story “Is this any way to save a species?” (*Nature* **436**, 14–16; 2005) implies that US taxpayers were not well served by the narrow funding of about \$120 million between 2001 and 2004 to promote the recovery of an endangered population of Steller sea lions. I disagree with this suggestion.

Alaska produces roughly half of the United States’ commercial fisheries landings by volume. The economic foundations for most coastal communities in the Gulf of Alaska and Bering Sea are marine-fisheries products, and the families that live in these communities would be hard pressed to make a living without these fisheries. From this perspective, an investment in research and management over a four-year period of less than 10% of the annual value of the fishery seems prudent. In any event, the legal impasse that had been reached between the National Oceanic and Atmospheric Administration (NOAA) and the environmental group bringing suit would have resulted in an estimated loss of annual revenue of between \$200 million and \$500 million.

Another important point is that the \$120 million supported not only research on Steller sea lions in Alaska, but a complex ecosystem-based research programme. The Congressional language referred to in your article directed NOAA to conduct research on a comprehensive set of factors that might have contributed to the decline in abundance of sea lions. These include climate change in the Gulf of Alaska and the Bering Sea, pollution, disease, deaths caused by human involvement (such as entanglement in

marine debris, poaching, vandalism and Alaska native subsistence harvests) and the impacts of predators on sea lions and of commercial fisheries on Steller sea lions’ prey fishes. Research on all these factors was carried out at varying levels of sophistication. In addition, NOAA used these funds to support various management activities relevant to the two populations of Steller sea lions in Alaska.

Although it is too early to conclude that a recovery is under way, and too soon to definitively assign cause and effect, the conservation measures developed by the North Pacific Fishery Management Council and implemented by NOAA are associated with the first increase in abundance in the western population of Steller sea lions in more than 30 years. If this trend continues for two more biennial censuses, this population will indeed be recovering.

Of course, \$120 million is a great deal of money. Doing research in the difficult field conditions of Alaskan waters is expensive and time-consuming, particularly given the enormous distances covered in the sea lions’ geographical range. Although many environmental issues remain to be addressed, the example in Alaska shows that it is possible to achieve sustainable fishing as well as protection of endangered species.

William Hogarth

NOAA Fisheries Service, National Oceanic
and Atmospheric Administration,
US Department of Commerce, 1315 East-West
Highway, Silver Spring, Maryland 20910, USA

A woman's place in Nature

SIR — It is astonishing to find, in the celebrated new format of *Nature*, a prominent transcription of the 1869 *Nature* mission statement: “...to aid scientific men themselves...”.

This assertion harks back to archaic times when women had not achieved recognition in science. It does not correspond to the open spirit that has, for a long time, characterized *Nature* and it should be modified in line with the present, more equitable, times.

Fabio Salamanca-Buentello*, **Leonor Buentello-Malo†**, **Fabio Salamanca-Gómez‡**

*Joint Centre for Bioethics and Canadian Program on Genomics and Global Health, University of Toronto, Toronto, Ontario M5G 1L4, Canada
†Institute of Research in Anthropology, UNAM, Ciudad Universitaria, México, DF 04510, Mexico
‡Unit of Medical Research in Human Genetics, Apartado postal 12-951, México, DF 03020, Mexico

Contributions to Correspondence may be submitted to corres@nature.com. They should be no longer than 500 words, and ideally shorter. Published contributions are edited.

BOOKS & ARTS

Growth factors

Putting the ideas of Russian agronomist Trofim Lysenko into political and scientific context.

The Lysenko Effect: The Politics of Science

by Nils Roll-Hansen

Humanity Books: 2005. 335 pp. \$25

Garland E. Allen

The theories on plant breeding, hybridization and selection propounded by Trofim D. Lysenko from the mid-1920s became a *cause célèbre* in international scientific circles. In 1948, his views were enshrined as the official basis for the Soviet Union's agricultural policy. As the Cold War was reaching its height, the West held up the Lysenko case as a perversion of the liberal view of science as an autonomous enterprise into a 'socialist' view of science controlled by authoritarian and repressive political figures.

A cornerstone of Lysenko's plant-breeding programme was 'neo-lamarckism', the idea that changes to an organism during its lifetime can be transmitted to its offspring. This notion was highly controversial both inside and outside the Soviet Union, and by the mid-1930s it had been rejected by most Western biologists. How a whole agricultural programme came to be based on theories rejected by the rest of the world's biologists is a question that has intrigued historians ever since.

The problem with the traditional historiography of the Lysenko case is the view that Lysenko's science was bogus quackery from the beginning; that it was imposed from the top by Stalinist terror; and that biologists and agronomists in the Soviet Union were 'eliminated' if they voiced any criticism.

In *The Lysenko Effect*, Nils Roll-Hansen does not apologize for some of the worst transgressions of 'lysenkoism', but after a close reading of much of the published primary literature, as well as the extensive use of newly opened archives, he is able to paint a much more complex picture. He has achieved what no previous author has attempted: to take seriously the science propounded by Lysenko, and to understand how the debates within Soviet scientific and agricultural circles were framed in the light of the prevailing biological theory. The result is a refreshing look at a familiar but traditionally misunderstood episode in the history of science that has relevance for discussions about the organization of science, science policy and the relationship between scientific theory and technological practice today.

According to Roll-Hansen, to understand



Taking root: the policies of Trofim Lysenko (left) were officially approved by the Soviet Union in 1948.

Lysenko's rise to prominence it is necessary to separate his work in plant physiology from his later anti-mendelian and neo-lamarckian theories. Roll-Hansen provides a detailed examination of Lysenko's background and early research, set in the context of the history of work in Russian plant physiology. The factors stimulating germination and early flowering in a wide variety of crop species had attracted considerable research interest in Russia and Germany in the late nineteenth and early twentieth centuries. One of the key debates that had emerged by the 1920s was between proponents of day length (photoperiodism) and exposure to cold temperature (vernalization) as the major factors stimulating early germination and, more importantly, flowering.

The 1917 Bolshevik revolution and subsequent civil war disrupted agriculture, and by 1921, food shortages were acute. The critical problem for Soviet agriculture was to increase yield, both by learning how to manipulate environmental conditions and by developing genetic strains that could flower early and thus produce two crops in a season. Cold treatment had long been known to affect flowering time, but it was not clear precisely how best to use it;

moreover, what worked for one strain in one locality did not necessarily work elsewhere. For a country where much of the arable land lies above the latitude of Minnesota, these are not inconsequential issues. It was in the context of this debate in the early 1920s that the young Lysenko made his scientific debut.

Despite coming from a peasant background and being largely self-educated, Lysenko graduated from the Kiev Agricultural Institute in 1925. His early papers on plant physiology, particularly vernalization, were not groundbreaking, but they attracted the attention of Nikolai Vavilov, the leading figure in Russian plant biology and agriculture at the time. Vavilov became a staunch supporter of Lysenko's work until the late 1930s. Lysenko eventually linked vernalization and selection to create genetically stable lines of early flowering varieties, and to bring flowering times into synchrony so he could make hybrids between strains from different areas with different flowering times. It was when he sought to convert one strain into another by 'education' — that is, by repeated exposure to low temperatures so that the plant's acquired adaptation to cold eventually becomes inherited — that his

Particularly problematic for Soviet geneticists was the association of mendelian theory in the West with opposition to the darwinian theory of natural selection and with eugenics. Darwinism, with its emphasis on variation and selection, had received strong support in Russia since the nineteenth century. Mendelian genetics, which emphasized the stability of the gene, and Wilhelm Johannsen's pure-line experiments, which showed the limits of selection, were seen as contradictions to Darwin

JGS INC.

to comment on science, but also represents a form of research and knowledge production in its own right, though one belonging to a radically different epistemological tradition. Science, engineering and technology shape the world in which we live, but Ede shows us the role played by art in the ever more complex interplay of forces between science, technology and society. It is the artist who asks about the social effects of scientific developments and challenges the changing scientific concepts of life itself — these questions become ever more urgent with every scientific advance. Moving beyond the postulated dichotomy of the objective sciences and the subjective arts, the impressive abundance of contemporary artworks cited by Ede shows us that art is no longer limited to the production of beautiful artefacts, but has established its role as a legitimate form of knowledge production in its own right.

The engagement of art with science ranges from artists' iconological handling of scientific imaging to research projects executed as artistic endeavours by artists working in the laboratory. An example of the former is the work of Neal White, one-time artist in residence at the human genome project at Hinxton near Cambridge, UK; an example of the latter is the work of the Portuguese artist Marta de Menezes, who uses the laboratory technique of microcautery to modify the patterning of butterfly wings. Such artistic interventions in genetics and biological forms have made possible new means of artistic expression and art forms.

Some of the insights that art provides into the latest hot topics in science, such as cloning or the production of artificial chimaeras, have been extensively addressed by Suzanne Anker and Dorothy Nelkin in their book *The Molecular Gaze* (Cold Spring Harbor Laboratory Press, 2004). The use of biological materials by artists ranges from tissue engineering to stem-cell technologies and even transgenic animals, a phenomenon that raises ethical questions with regard to both scientific and artistic endeavours.

New directions in research, such as those offered by neurobiology and studies of consciousness, provide greater insight into the working of the mind, and molecular biology continues to provide us with a better understanding of the structure of the living world. Their scientific explanations of the structures and processes of body and mind challenge our conception and understanding of what we call 'human nature'. But individuality and self must be more than mere bundles of impulses, sensations and chemical processes.

Through the use of video endoscopies in her 1994 work *Corps étranger*, artist Mona Hatoum blurs the boundaries between the inner and the outer, allowing the viewer to participate in her own stream-of-consciousness and somatic experiences. Collaborating with neuroscientists, artist Annie Cattrell

uses functional magnetic resonance imaging (fMRI) data to create three-dimensional, amber-coloured works in resin that are then embedded in solid, clear, resin cubes or 'brain boxes' (see *Nature* 424, 18, 2002). Whether working with brain scans produced by advanced imaging processes or simply with traditional media, the focus of the artistic process is increasingly the diversity of human experience, something that often does not lend itself to portrayal using standard scientific procedures.

From the end of the nineteenth century onward, art has increasingly turned away from the classical quest for order, and has struggled on many levels with the disintegration of a uniform world view and a coherent conception of humanity. Scientific images today offer us amazing insights, but they must still be viewed as historical snapshots. Although modern science can provide us with ever more

detailed pictures of the inner workings of our bodies and of the living world, the influence of such images on our understanding of the nature of humanity remains an issue for social discourse.

Ede not only offers an overview of contemporary art practices, but also examines their philosophical background. Additionally, the artworks discussed are accompanied by extensive examples of contemporary science and research, providing further insight into the newest scientific developments. The book is an excellent contribution to the literature in the field of art and science, and provides a perspective that reaches far beyond the usual approaches to the relationship between, and intersection of, art and science. ■

Ingeborg Reichle is an art historian and theorist at the Berlin-Brandenburg Academy of Sciences and Humanities, Jägerstrasse 22/23, Berlin 10117, Germany.

Focusing on the stars

Stargazer: The Life and Times of the Telescope

Fred Watson

Da Capo Press: 2005. 342 pp. \$24.95

Robert H. van Gent

The telescope symbolizes the science of astronomy and had a pivotal role in the development of early modern science. Several excellent historical studies have already been written about it, notably André Danjon and André Couder's *Lunettes et Télescopes* in 1935, Henry King's *The History of the Telescope* in 1955, and Rolf Riekher's *Fernrohre und ihre Meister* in 1990.

As we approach the fourth centennial of the invention of the telescope, the Australian astronomer Fred Watson presents a well-written, up-to-date history of the invention and development of the telescope and its impact on astronomy. As the astronomer-in-charge at the Anglo-Australian Observatory at Coonabarabran in New South Wales, Watson is well equipped to write such a history. He tells the fascinating story of the invention of simple telescopes by Dutch spectacle-makers in the early years of the seventeenth century, and recounts their evolution into the modern telescopes of today.

The reliably documented history of the telescope begins in

September 1608, when the Middelburg spectacle-maker Hans Lippershey submitted a request for a patent on his invention. But Watson also dwells on the (unlikely) possibility that primitive telescopes had been known to the ancient Mesopotamians, and to Hellenistic and medieval scholars in Europe.

A stronger case can be made for the so-called Elizabethan telescope, an early reflecting telescope described in the works of the Englishmen Thomas Digges and William Bourne, but the evidence suggests that it existed only in the minds of those who wrote on it, rather than in reality.

Although the inventors of the refracting telescope were Dutchmen, it fell to Galileo, Thomas Harriot and Simon Marius to use it to observe the Universe. Their observations soon provided proof for the Sun-centred copernican model of the heavens. Christiaan Huygens, Johannes Hevelius and William Herschel all made key discoveries using telescopes that they had designed and built themselves.

After the early successes, Watson goes on to describe the subsequent improvements of the refracting and reflecting telescopes, with considerable attention to detail. In

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Star performer: Galileo's telescope from the early seventeenth century.

the final chapters, he discusses the great Earth- and space-based telescopes of the twentieth century that revolutionized our understanding of the origin and fate of the cosmos, and he offers a peek into the near future at a new generation of super telescopes that will probe even further into space.

Throughout the text there are useful diagrams and illustrations that ably illustrate the

various lens and mirror configurations that have been designed during the telescope's first 400 years. There are also detailed notes, a glossary of astronomical and optical terms, and a bibliography.

Inevitably, there are a few minor errors. For example, Ptolemy's *Almagest*, an astronomy manual, should be dated to the middle of the second century AD, not the first century AD.

Watson's book is a welcome addition to the literature on the history of the telescope, and can be recommended to any reader with an interest in the history of science and instrumental technology.

Robert H. van Gent is at the Institute for the History and Foundations of Science, Utrecht University, Utrecht NL-3508 TA, The Netherlands.

Marine undercurrents

An installation by Ellen Gallagher builds on some little-known work by Sigmund Freud.

Colin Martin

The house in Hampstead, London, where Sigmund Freud lived during the last year of his life, after leaving Vienna in 1938, has been a museum since 1986. His youngest daughter Anna, who died in 1982, left the study and library exactly as Freud used them, crammed with his furniture, books and extensive collection of antiquities. The house has a palpable sense of intellectual continuity, which has inspired contemporary artists to engage with Freud's ideas and create work for display there.

American artist Ellen Gallagher's current installation, *Ichthyosaurus*, cites early work by Freud that he rarely mentioned in his later years. Although best known as the founder of psychoanalysis, Freud also worked as a neurologist and anatomist between 1876 and 1896. He began his scientific career assisting the physiologist Ernst Wilhelm von Brücke in Vienna, studying the spinal nerve ganglia of a primitive fish, the lamprey (*Petromyzon*), to gain insight into the evolution of nervous systems.

Fifteen of Freud's drawings of *Petromyzon* from 1876 and 1877 (like the example shown on the right) are displayed for the first time, in conjunction with Gallagher's installation. Her interventions in his library and study are discreet, and their extent becomes apparent only after looking carefully into the cordoned-off rooms. Immediately obvious are two 16-mm films made in collaboration with Edgar Cleijne, *Ichthyosaurus* 2005, which are projected onto the walls. Although they initially appear to be family home movies, they are cinematic riffs on Gallagher's marine theme: one shows flailing



has hung three mixed-media works from her 'Watery Ecstatic' series, begun in 2001, on his walls.

Fortunately, it is possible to view two further works located in the hall more closely and appreciate her skilful use of various techniques. These include intricately cut and coloured paper to form fish scales in *Watery Ecstatic* 2005 and cut layers of uncoloured paper to form fossil-like imprints of marine species in *Watery Ecstatic* (22 078 N, 159 322 W) 2005.

Although her work is beautiful, it is unclear how Gallagher's appropriation of marine imagery provides more than a cursory nod towards Freud's early scientific work. His intention was simply to make

drawings that presented cellular structure clearly, whereas she attempts to present multilayered personal and cultural meanings in her work.

Freud's influence appears more marked in Gallagher's two playful collages inspired by surrealism, the artistic movement that evolved from his psychoanalytical theories. In her photomontage *Odalisque* 2005 (shown here, above left), a seductively reclining Gallagher fails to divert Freud from his drawing board. In *Abu Simbel* 2005, she reworks a copy of the photogravure of an Egyptian temple that used to hang above Freud's couch in Vienna, adding a marine-like spacecraft complete with blue fur tentacles, which zaps rays at the colossal statues of Rameses II.

Ichthyosaurus can be seen at the Freud Museum (www.freud.org.uk) in London until 11 September.

Colin Martin is a London-based writer.



strands of seaweed, and the other spotlights various marine specimens that are difficult to discern from a distance.

Gallagher has also placed three glass specimen jars containing fanciful marine creatures among Freud's own objects, and

NEWS & VIEWS

R. HAMILTON SMITH/CORBIS



NEUROSCIENCE

Finding the missing fundamental

Robert J. Zatorre

The whole orchestra tunes up to an A note from the oboe — but how do our brains tell that all the different sounds are the same pitch? The discovery of pitch-sensitive neurons provides some clues.

Although Maurice Ravel reportedly came to regret ever having written *Bolero*, it has become a popular staple of the orchestral repertoire. It relies entirely on a single theme, repeated over and over (and over) by different combinations of instruments. Artistic merit aside, the piece raises an interesting question: how do we effortlessly recognize the same melody played by different instruments even though the acoustical structure of the sound reaching our ears varies with the instrument? Bendor and Wang (page 1161 of this issue)¹ have found neurons that figure out what the pitch of a sound is even when they are presented with physically different signals, giving hints to how we come to perceive pitch as a unified entity.

Psychologists are intrigued by the problem of perceptual constancy: essentially, how do we perceive the environment as remaining stable despite huge variability in the inputs reaching our senses? This general question is especially puzzling in the case of pitch, because we have known since the nineteenth century² that the pitch of a sound typically corresponds to its fundamental vibrational frequency — even if that frequency is physically absent in the sound reaching our ears. Sounds that have pitch arise from objects that vibrate in a periodic manner, such as columns of air in pipes or the vocal cords (as opposed to aperiodic sounds like wind or rushing water). As Pythagoras knew, if you pluck a string, it will vibrate in its entire extent, as well as in halves, thirds and so on, and each of those vibrational

modes will result in a separate harmonic frequency. Yet we usually perceive the pitch as corresponding to the lowest of these, which is the fundamental³. For a simple demonstration of the 'missing fundamental' effect, pick up a phone. Most telephone lines cut off the lower frequencies, resulting in a slightly tinny sound, yet the fundamental pitch does not change; a male voice does not sound like Mickey Mouse. The brain seems to figure out the missing pitch.

Bendor and Wang¹ studied the auditory cortex (the region of the brain that enables perception of sound) in the marmoset monkey. They show that there are neurons in this region that respond in essentially the same way to a variety of sounds that all have the same fundamental but do not share any frequencies. For example, a neuron that responds to 200 hertz also responds to the combination of 800, 1,000, and 1,200 hertz because all correspond to the same fundamental. This effect is unusual because neurons usually respond only within their receptive field, which is typically a narrow range of frequencies. The marmoset neurons, however, responded not only to frequencies in their receptive fields, but also when there was no frequency within the receptive field but the other frequencies in the stimulus were harmonically related to the missing one. This property makes psychologists happy, because it provides evidence (if not yet a mechanism) for perceptual constancy. These neurons respond to an

abstract property — pitch — derived from, but not identical to, physical sound features. Presumably, therefore, it is thanks to such neurons that we can follow a tune as the instruments change.

One might wonder why marmosets need such a system, given that they don't spend much time listening to iPods. But periodic sounds are important in the natural environment because they are almost exclusively produced by other animals, and so pitch is a good cue to segregate these sounds from background noise⁴. Marmosets are highly vocal creatures, and the development of pitch-sensitive neurons would also be central to communication. From an evolutionary perspective, these abilities could be seen as precursors to human pitch perception, which has led to our unique development of music and is similarly crucial for speech.

The location of the pitch-sensitive cells lateral to the primary auditory cortex, as described by Bendor and Wang, is compatible with studies of the human brain. In human patients, damage to areas analogous to the marmoset pitch-sensitive regions produce specific deficits in perceiving missing fundamental pitch⁵. Moreover, neuroimaging studies in humans demonstrate pitch sensitivity in roughly the same location^{6,7}. The human studies typically show specialization for pitch in the right auditory cortex, however. Bendor and Wang do not address this issue, as only a single hemisphere was probed in each of three

monkeys. It will be of interest to determine whether lateralization is present in other species (as others suggest⁸), and is therefore related to basic properties of sound processing, or whether it is uniquely human and thus might be a consequence of the development of language.

Now that we know that there are pitch-sensitive neural units, we have to discover how they work. Sound undergoes many transformations before it gets to the auditory cortex, resulting from the biophysical properties of the cochlea and the many neuronal junctions between cochlea and cortex. We do not yet know precisely how periodic, temporal information available in a stimulus is integrated with the spectral information (or individual harmonics) that is also extracted by the system. We also do not know much about the inputs to the neurons described by Bendor and Wang. Do they come in a hierarchical arrangement from other simpler cells in the auditory cortex? Or do they also receive inputs from subcortical structures such as the thalamus? Perhaps

top-down influences from centres associated with complex functions in frontal or parietal lobes are also significant. This last point is relevant, because one technical advantage of this work is that the animals tested were awake rather than anaesthetized, meaning that attentional and other cognitive factors could have a role. The animals were not trained or behaving, however, so it is difficult to know the significance of the stimuli for them. Understanding the interaction between basic perceptual systems and their modulation by higher-order mechanisms will require more attention to these factors. Another interesting question is whether these neuronal properties are somehow hard-wired, or whether they are a consequence of the animals' environmental experience with periodic sounds, which contain harmonically related frequencies.

Ian Whitfield⁹ noted that the problem of perception is not to determine that two events are different, which is actually fairly trivial, but rather that events that might seem to be different are actually the same. It is the job

of the cortex, he argued, to perform the computations needed to extract invariances despite the different inputs that the environment may provide. The present study¹, and those that will no doubt follow, will lead to a more profound understanding of this fundamental problem.

Robert J. Zatorre is at the Montreal Neurological Institute, McGill University, 3801 University Street, Montreal, Quebec H3A 2B4, Canada. e-mail: robert.zatorre@mcgill.ca

1. Bendor, D. & Wang, Q. *Nature* **436**, 1161–1165 (2005).
2. von Helmholtz, H. *On the Sensations of Tone as a Physiological Basis of the Theory of Music* (1863) (trans. Ellis, A.; Dover Publications, New York, 1954).
3. Schouten, J. *Koninklijke Nederl. Akad. Wetenschappen Proc.* **41**, 1086–1093 (1938).
4. Bregman, A. *Auditory Scene Analysis* (MIT Press, Cambridge, Mass., 1990).
5. Zatorre, R. J. *J. Acoust. Soc. Am.* **84**, 566–572 (1988).
6. Patterson, R. D., Uppenkamp, S., Johnsrude, I. S. & Griffiths, T. D. *Neuron* **36**, 767–776 (2002).
7. Zatorre, R. J. & Belin, P. *Cerebral Cortex* **11**, 946–953 (2001).
8. Wetzel, W., Ohl, F., Wagner, T. & Scheich, H. *Neurosci. Lett.* **252**, 115–118 (1998).
9. Whitfield, I. in *Cerebral Cortex* (eds Peters, A. & Jones, E.) 329–351 (Plenum, New York, 1985).

BIOLOGICAL CHEMISTRY

Just add chlorine

Nathan A. Schnarr and Chaitan Khosla

Nature provides lessons about developing 'green chemistry' in seemingly out-of-the-way places. One such lesson comes from an enzymatic step in the production of a leaf toxin by a bacterium.

As they describe on page 1191 of this issue¹, a group of researchers led by Christopher Walsh has identified how chlorine is attached enzymatically to an intermediate during the formation of a natural product. This is not surprising in itself — the significance lies in the unreactive

nature of the carbon centre concerned.

Many natural products require halogens (chlorine, bromine or iodine) to be strategically placed onto organic molecules at unreactive carbon centres. Halogenation is essential to the biological activity and chemical reactivity of

such products (Figs 1a–c), and often serves to generate versatile molecular building blocks for synthetic organic chemists. Ideally, these syntheses would use alkanes — unreactive carbon chains — as their starting materials. These are usually readily available and relatively cheap as they are the main components of oil. Unfortunately, traditional methods for incorporating halogens into alkanes often require environmentally unfriendly reagents and suffer from poor control of specificity (Fig. 2a). In contrast, natural enzymes are benign and do the same job with extra-ordinary specificity, but little is known of the mechanisms of these enzymatic halogenations. Now, Walsh and colleagues¹ have discovered that halogenation of an unreactive carbon centre can be catalysed by a halogenase enzyme, called CmaB, that is α -ketoglutarate dependent and contains non-haem iron.

Similar catalysts are known to be involved in oxygenation chemistry carried out by the hydroxylase family of enzymes. Hydroxylases insert oxygen into a carbon–hydrogen bond, an analogous process to halogenation, and have received much attention because of their extraordinary specificity and versatility. In nature, several different types of hydroxylase catalyse such transformations, depending on the substrate. In general, the more reactive substrates require the less reactive enzymes. For example, hydroxylation of highly reactive *p*-hydroxybenzoic acid is readily accomplished by a flavin-dependent hydroxylase. In contrast, hydroxylation of relatively unreactive substrates, such as the amino acids proline or lysine, requires significantly stronger α -ketoglutarate-dependent enzymes containing non-haem iron². In addition to these two extreme cases, a variety of alternative hydroxylases has

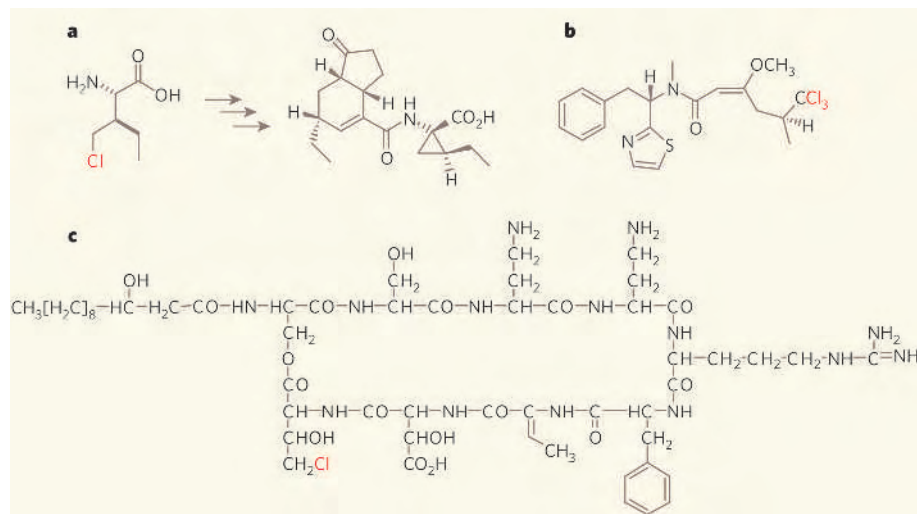


Figure 1 | Chlorine in natural-product synthesis. **a**, Coronatine, a leaf toxin. The chlorinated intermediate (left) goes through several further reactions before coronatine, which does not itself include chlorine, is produced. **b**, Barbamide, a molluscicide. **c**, Syringomycin, an antibiotic. Biosynthesis of all three products involves enzymatic chlorination.

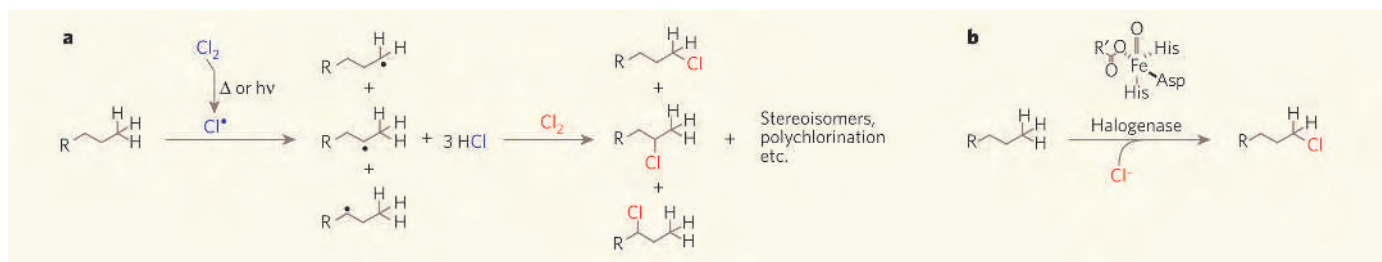


Figure 2 | Chlorinating unreactive organic compounds. **a**, Using purely chemical synthesis without an enzyme, alkanes are chlorinated using highly reactive radical species (black dots). This requires several steps, and it is difficult to control which particular carbon is modified and included in the final stereochemical form of the product. **b**, The reaction is much easier and the products more controllable with a dedicated halogenase, as described by Walsh and colleagues¹. A chloride anion is oxidized by the oxo-iron species from the halogenase active site for site-specific chlorination of the substrate.

been discovered that act primarily on substrates with intermediate reactivity. By analogy with the hydroxylases, Walsh and colleagues now speculate that natural systems may employ a similar, tunable strategy for halogenation.

In this model, more reactive aromatic substrates rely on gentler, FADH₂-derived agents — and this is indeed the case for chlorination of tryptophan during the biosynthesis of the antitumour agent rebeccamycin³. Unreactive carbon centres, in contrast, require a more vigorous agent. In the reactions investigated by Walsh and colleagues, the necessary oxidation of a chloride anion to add chlorine to an alkane is evidently carried out by a highly reactive oxo-iron species from the halogenase active site (Fig. 2b). This species presumably breaks an unactivated carbon–hydrogen bond, leading to a high-energy radical species that can be trapped by chlorine.

Analogous enzymes with an apparently similar mode of action are already emerging from other biosynthetic systems. For example, Walsh and colleagues⁴ have also identified a putative halogenase in the biosynthetic pathway for the antifungal antibiotic syringomycin (Fig. 1c). This enzyme is presumably responsible for introducing chlorine into the natural product. Similarly, two relatives of CmaB (BarB1 and BarB2) are found in the biosynthetic pathway for barbamide, a potent molluscicide that contains a medicinally interesting trichloromethyl group (Fig. 1b).

Generation of high-energy radical intermediates has emerged as the common thread among many oxo-iron catalysts, whose functions range from natural product biosynthesis to post-translational protein modification and even repair of RNA or DNA. The addition of halogenation to this impressive array of activities illustrates yet another exciting way to generate useful chemical intermediates from comparatively unreactive precursors. Elucidation of the precise mechanism for this transformation will undoubtedly pave the way for novel organometallic halogenation catalysts for chemical synthesis.

Nathan A. Schnarr and Chaitan Khosla are in the Departments of Chemistry and Chemical Engineering, Stanford University, Stanford, California 94305-5025, USA.
e-mail: khosla@stanford.edu

1. Vaillancourt, F. H., Yeh, E., Vosburg, D. A., O'Connor, S. E. & Walsh, C. T. *Nature* **436**, 1191–1194 (2005).
2. Hausinger, R. P. *Crit. Rev. Biochem. Mol. Biol.* **39**, 21–68 (2004).

3. Yeh, E., Garneau, S. & Walsh, C. T. *Proc. Natl Acad. Sci. USA* **102**, 3960–3965 (2005).
4. Vaillancourt, F. H., Yin, J. & Walsh, C. T. *Proc. Natl Acad. Sci. USA* **102**, 10111–10116 (2005).

EARTH SCIENCE

Helium not in store

William M. White

The ratio of helium isotopes in some oceanic volcanoes seemed to demand a reservoir of virgin primordial gas in the Earth's mantle. In fact, that might not be necessary — a relief for other geophysical models.

In the 4.5 billion years of its existence, Earth has been continually losing helium through the degassing of rocks from its interior as they melt during volcanic processes. That there is any helium left on Earth at all is largely owing to its replenishment in the interior through radioactive α -decay, principally of the heavy elements thorium and uranium — the α -particle emitted in α -decay is a helium nucleus. But α -decay creates only the heavier helium isotope, ^4He . Any trace of the lighter isotope, ^3He , on present-day Earth is primordial, dating from the planet's formation.

Admittedly, Earth does not have much ^3He : in interior rocks there is only one atom for every 100,000 atoms of ^4He . But the $^3\text{He}/^4\text{He}$ ratio is even smaller in the atmosphere — typically eight or nine times lower, but occasionally up to 40 times lower. This imbalance would seem to imply that Earth still retains substantial amounts of primordial helium trapped in its interior. But where? Is this gas largely confined to a single deep reservoir that has remained undisturbed by volcanic activity all this time? Or is it more uniformly dispersed, with degassing just less efficient than we had thought? On page 1107 of this issue, Class and Goldstein¹ argue strongly for the latter.

Volcanoes allow a glimpse into the evolution of processes in Earth's interior. Isotopic analysis of radiogenic elements implies that material from Earth's mantle (the layer between crust and core) that is disgorged as volcanic lava has generally been melted before, at least partially. Helium, a noble gas, is not

chemically bound in minerals, so should escape by degassing whenever melting occurs, first to the surface, and from there into space. And although ^3He is replaced by radioactive decay, ^3He is not; the higher the $^3\text{He}/^4\text{He}$ ratio, therefore, the less melting and degassing has taken place.

The existence of rock with high $^3\text{He}/^4\text{He}$ ratios has led to the notion that there is a reservoir of rock, most reasonably in the deepest mantle, that has escaped melting and degassing and retains much or all of its original helium. This idea is supported by the highest $^3\text{He}/^4\text{He}$ ratios being found in the lavas of oceanic island volcanoes, such as those on Hawaii and in Iceland (Fig. 1). These volcanoes are thought to be produced by convection plumes that carry hot rock from the deep mantle^{2,3}. Yet if a deep reservoir of rock exists in its primordial state, it must be isolated from the convection that affects the rest of the mantle and drives plate tectonics. Seismic imaging of Earth's interior has, however, consistently failed to find evidence of any layering in the deep mantle, and implies instead that the whole mantle is involved in convection⁴.

The observed ratios of helium isotopes are therefore problematic. They seem to require layered convection that seismologists cannot detect and geodynamicists cannot reproduce in their models. They seem to require a primordial deep mantle, whereas the isotopic ratios of other radiogenic elements indicate that all of Earth's interior has been affected by earlier volcanic activity.

Class and Goldstein¹ attempt to reconcile

D. PEEBLES/CORBIS

these incompatibilities. They investigated the strontium (Sr), neodymium (Nd) and lead (Pb) isotopic ratios of the oceanic island basalts with the highest $^3\text{He}/^4\text{He}$ ratios and show that these rocks are derived from mantle that is relatively depleted in so-called incompatible elements — elements that are not readily accommodated in mantle minerals and are easily extracted by partial melting.

Against expectation, therefore, it is rocks that have most obviously been melted and undergone degassing that have the highest $^3\text{He}/^4\text{He}$ ratios. Lower $^3\text{He}/^4\text{He}$ ratios, indicating a high degree of degassing, are found in those oceanic island basalts whose other isotope ratios are closest to the expected primordial values. The similarity of these rocks to primordial mantle could be coincidence, the result of incompatible-element depletion by melting and subsequent re-enrichment, perhaps by addition of material subducted from Earth's crust.

In the second part of their paper¹, Class and Goldstein report model calculations that show that $^3\text{He}/^4\text{He}$ ratios as high as those observed in some oceanic basalts could be preserved in the mantle despite extensive melting, volcanism and degassing. The degree to which a model reflects reality always depends on its guiding assumptions: Class and Goldstein assume, for example, that noble gases are not extracted with near-perfect efficiency during melting, but behave like highly incompatible elements, which allows some primordial ^3He to be retained. Further-



Figure 1 | Erupting evidence — Kilauea volcano on Hawaii.

more, surface tension dictates that you can no more get all the melt out of a partially molten rock than you can get all the water out of a kitchen sponge. If some of the melt remains in the rock and eventually resolidifies, some of the helium will remain as well. Just how much remains is difficult to judge with our present knowledge; this model may stand or fall with further research on the chemical behaviour of noble gases during melting and on the physics of partially molten rock.

The other central assumption of Class and Goldstein's model is Earth's initial helium abundance: the greater this was, the higher the $^3\text{He}/^4\text{He}$ ratio will be now. The abundance of the noble-gas isotope xenon-129 in the

mantle, the decay product of now-extinct iodine-129, indicates that the Earth experienced catastrophic degassing very early in its history⁵ — quite possibly as a result of the collision that formed the Moon. This would have released much of the planet's primordial helium. How much was left behind no one knows. This helium abundance, the starting-point of Class and Goldstein's model, is therefore essentially an unconstrained parameter. We shall see how it stands up to scrutiny.

Although some may be reluctant to relegate 'primordial mantle' to the scientific graveyard quite yet, the case made by Class and Goldstein will be hard to rebut. Unsettled controversies remain: for example, there is still a need to maintain separate reservoirs in the Earth's interior to explain variations in other isotope ratios, although for much shorter times than the age of the Earth. This is difficult to reconcile with fairly strong geophysical evidence for convection involving the whole mantle that would destroy that separation. There is much still to learn about the structure and evolution of the Earth's deep interior. ■

William M. White is in the Department of Earth and Atmospheric Sciences, Cornell University, Ithaca, New York 14853, USA.

e-mail: white@geology.geo.cornell.edu

1. Class, C. & Goldstein, S. J. *Nature* **436**, 1107–1112 (2005).
2. Morgan, W. J. *Nature* **230**, 42–43 (1971).
3. Montelli, R. *et al.* *Science* **303**, 338–343 (2003).
4. Grand, S. P. *Phil. Trans. R. Soc. Lond. A* **360**, 2475–2491 (2002).
5. Staudacher, T. & Allègre, C. J. *Earth Planet. Sci. Lett.* **60**, 389–406 (1982).

BIOLOGICAL CHEMISTRY

Enzymes in focus

Romas Kazlauskas

The technique of directed evolution creates thousands of mutant enzymes from a single original. A new approach helps to search for variants that have an increased range of substrates — and can thus be used for organic synthesis.

Protein catalysts, or enzymes, are useful in organic synthesis largely because they can accept substrates other than their natural ones. Yet they can still distinguish subtle differences in shape between substrates — a characteristic known as stereoselectivity. The substrate range and specificity of an enzyme can be modified by protein engineering. In this case, mutants are created by changing the enzyme's component amino acids. Such mutant enzymes can be used, among other things, to synthesize pharmaceutical building-blocks. Writing in *Angewandte Chemie*, Reetz *et al.*¹ demonstrate a variation on recently developed enzyme-engineering methods^{2–6}. They mutate pairs of amino acids in the enzyme's substrate-binding site (active site) to create variants with an increased range of substrates

but that retain high stereoselectivity.

Initial efforts at enzyme engineering took a so-called rational-design approach. This involved using knowledge of enzyme structure and active sites, together with computer modelling, to predict precisely the mutations needed. Success was not only measured in terms of increased specificity, but also of stability, the ability to fold, and catalytic activity. Many of the early attempts were disappointing, as these interdependent properties are hard to predict. To increase their success rate, researchers developed techniques such as saturation mutagenesis, in which the effects of each of the 20 normal amino acids are tested at selected positions in the enzyme⁷.

The discovery of the polymerase chain reaction (PCR), which copies DNA strands

extremely fast, greatly simplified molecular-biological techniques. As a result, the emphasis in enzyme engineering shifted from using rational design to 'directed-evolution' tools that rely on random mutagenesis. Such techniques, for example error-prone PCR and 'gene-shuffling', involve randomly and repeatedly varying amino-acid residues throughout the enzyme. This creates enormous numbers of mutant enzymes. When screened for activity, however, typically only very few of these turn out to be useful.

Surprisingly, many of the mutations identified in directed-evolution experiments were found far from the active site — so far from it, in fact, that the mutated residue did not come directly in contact with the substrate⁸. So was it wrong to focus on the enzyme's substrate-binding site? Are distant mutations instead better at changing enzyme specificity?

The answer turns out to be no. Directed evolution discovers distant mutations not because they are more active, but because they are more common^{2,8,9}: there are simply more amino acids far from the active site than close to it. Thorough screening will still find the best mutants; but in practice, it is easier to generate large numbers of mutants than it is to screen them, and incomplete screening favours the more

common, distant mutations. So the emphasis in improving enzyme specificity has returned to the active site, where amino-acid mutations are likely to be more effective²⁻⁶. This approach is known as focused directed evolution.

Reetz and colleagues' new take on focused directed evolution¹ involves mutating pairs of amino acids in the active site; this process allows more extensive reshaping of the site than would a single mutation. The authors chose amino-acid pairs that pointed towards the substrate, and — simplifying the experiment considerably — chose only pairs that were close together in the linear sequence of amino acids that would form the folded enzyme unravelling. This meant that for each mutant pair, only one 'mutagenic primer' was required to initiate a PCR process and thereby create a library of mutants. Producing mutations in a pair of amino acids thus becomes no more difficult than in a single amino acid, although, as there are 20² different combinations of two amino acids, much more screening is required. Reetz *et al.* elegantly use secondary structure — how the enzyme folds on a local scale — to identify spatially interacting amino acids in the linear sequence (Fig. 1).

The authors tested their paired-mutation approach on a lipase — an enzyme that breaks down fats — from the bacterium *Pseudomonas aeruginosa*. They generated five mutant libraries, picking five different amino-acid pairs from around the active site. Each library contained 400 variants: the original, unmutated enzyme; 38 single mutants, in which one of the 19 'introduced' amino acids occurred at one of the pair of positions; and 361 (19 × 19) double mutants. The screening process identified eight mutants that typically showed 5–30 times greater activity than the unmutated enzyme. Five mutants were from one library, three from another.

Reetz *et al.* tested the enzyme variants on 11

substrates, and their preliminary results indicate that mutants showing increased activity also show good stereoselectivity. (It should be noted, however, that the mutants of this particular enzyme would not be sufficiently stable for applications in organic synthesis.) A previous directed-evolution study of this lipase¹⁰ identified mainly distant mutations that typically only doubled stereoselectivity. Interestingly, amino acid 162, which was changed in five of the eight best-performing mutants in Reetz and colleagues' study, was also identified as a key position in this previous work.

A surprising finding was that five of the eight best mutants were single mutants. For the 11 target substrates, a single mutant was most active in nine cases, whereas a double mutant (designated M16A, L17F) was most active in the other two. Researchers will need to decide whether the extra chance of success among the double mutants is worth 20 times more screening. One solution might be to use the earlier approach of choosing amino acids that are not adjacent in the linear amino-acid sequence, but are spatially adjacent when the enzyme folds into its natural shape⁴⁻⁶. However, this would require the slightly more complex use of two primers for mutagenesis.

It will always be possible to make many more mutants than can be screened, so strategies for choosing the mutations most likely to improve an enzyme's properties will continue to be important. If the structure (or a homology model) of an enzyme is available, focused directed evolution is currently the fastest approach to altering specificity. For other properties, such as thermal stability, targeting the whole enzyme remains the best approach⁹. ■

Romas Kazlauskas is in the Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, 1479 Gortner Avenue, Saint Paul, Minnesota 55108, USA.
e-mail: rjk@umn.edu

1. Reetz, M. T., Bocla, M., Carballeira, J. D., Zha, D. & Vogel, A. *Angew. Chem. Int. Edn* **44**, 4192–4196 (2005).
2. Park, S. *et al. Chem. Biol.* **12**, 45–54 (2005).
3. Hill, C. M., Li, W.-S., Thoden, J. B., Holden, H. M. & Raushel, F. M. *J. Am. Chem. Soc.* **125**, 8990–8991 (2003).
4. Whittle, E. & Shanklin, J. *J. Biol. Chem.* **276**, 21500–21505 (2001).
5. Koga, Y., Kato, K., Nakano, H. & Yamane, T. *J. Mol. Biol.* **331**, 585–592 (2003).
6. Li, Q.-S. *et al. Biochim. Biophys. Acta* **1545**, 114–121 (2001).
7. Wells, J. A., Vasser, M. & Powers, D. B. *Gene* **34**, 315–323 (1985).
8. Horsman, G. P. *et al. Chem. Eur. J.* **9**, 1933–1939 (2003).
9. Morley, K. L. & Kazlauskas, R. J. *Trends Biotechnol.* **23**, 231–237 (2005).
10. Liebeton, K. *et al. Chem. Biol.* **7**, 709–715 (2000).

EVOLUTION

A treasure trove of motors

Margaret A. Titus

The myosins are a superfamily of protein motors. Analysis of their sequences in a wide range of organisms reveals an unexpected variety of domains, and provides insights into the nature of the earliest eukaryotes.

Motor proteins use chemical energy, for example from ATP, to generate unidirectional movement along a filamentous track. How a group of proteins acquired and then varied this property to generate a range of movements as evolution proceeded is a fascinating problem in biology. Answers are within reach because of the availability of genome sequences from a diverse cadre of organisms representing various evolutionary groups. This allows in-depth comparative analyses of the sequences of protein families and the incorporation of these data into models of evolution. Richards and Cavalier-Smith (page 1113 of this issue)¹ have performed a comprehensive analysis of the myosin superfamily of motor proteins across a wide sample of eukaryotes (organisms whose cells have nuclei, including plants and animals). The results provide insights into how myosins evolved and into the nature of the earliest common ancestor — the eukaryotic cell.

The myosins are a diverse group of motor proteins that move along the actin filaments that form a major component of the cell's

internal scaffolding. Myosins are best known for powering muscle contraction, cell migration and cytokinesis (the separation of two daughter cells during cell division)^{2,3}. These proteins typically consist of an amino-terminal motor domain that binds to the actin track and catalyses nucleotide hydrolysis — the reaction used to harvest energy from ATP — and a carboxy-terminal tail region. This directs the motor domain to targets within the cell, either binding to a cargo that is to be transported around the cell or anchoring the myosin to a particular site^{2,3}.

Previous comparisons of myosin motor domains identified 18 distinct classes, each associated with specific tail domains, and revealed that the motor and tail appear to have evolved in synchrony⁴⁻⁶. Richards and Cavalier-Smith used the highly evolutionarily conserved sequence of the core myosin motor domain to search for myosin genes in genomes from five major taxonomic groups: the amoebae; the opisthokonts (which include fungi and animals); the excavates (which include those protozoans with flagella — whip-like

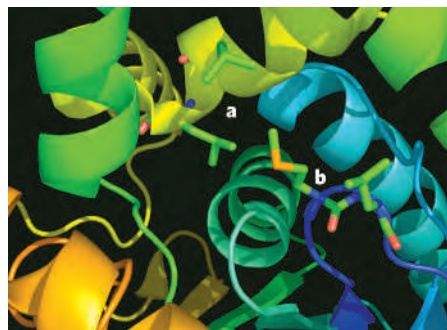


Figure 1 | View of the enzyme active site used by Reetz *et al.*¹ for pair mutation. The local secondary structure of an enzyme determines how far apart in the linear amino-acid sequence two spatially adjacent amino acids are. **a**, In a so-called 3₁₀ helix (light green), the two spatially interacting amino acids are three amino acids apart (the green stick-like protrusions are the side chains, while the blue and red tips are the main protein chain, which is shown only as a ribbon for the other amino acids). **b**, By contrast, in a loop (dark blue) the two amino acids are adjacent.



50 YEARS AGO

There is something very depressing about contemporary biological journals. Paper after paper records observations or experiments, analyses them cautiously, and in a timid and tentative way compares them with previous observations and experiments on the same theme. That is about all: only rarely does the writer disclose how (in his view) his work is related to the broad panorama of biology. There are doubtless sufficient reasons for these omissions: many writers of papers undertake the research they describe for no other reason than that their supervisors 'put them on to it', and many editors of journals consider contemplation out of place in science and do not encourage authors to indulge in it...How refreshing it is, for example, to hear that the choice of a subject for research involves the "art of rejection", and to be told that this art can be compared with the art of the Chinese in designing the empty spaces in their pictures. It is refreshing, too, to be reminded...that the very observations one makes, and *a fortiori*, one's interpretation of them, are limited by the Zeitgeist and by unconscious philosophical assumptions derived from Spinoza.

From *Nature* 27 August 1955.

100 YEARS AGO

A somewhat lamentable aspect of modern science is the vast array of unorganized facts which are awaiting coordination; this is too often because they have been amassed without any definite idea of the purpose which they may serve; consequently it may happen that laborious observations belonging to one science may fail to attract the regard of a neighbouring science merely for want of the mutual acquaintance which would make them serviceable to each other; and in these days of exclusive specialisation the introduction which might lead to a happy union is, perhaps, not brought about for years.

From *Nature* 24 August 1905.

tails that propel them); plants; and chromoalveolates (including dinoflagellates and the apicomplexan parasites such as the malaria-causing *Plasmodium*). The complete sequences of these myosin proteins were then used to find protein domains present in the myosin tails and at the extreme N-termini of some of the molecules.

The analysis uncovered a rich variety of myosins throughout the eukaryotes, extending the catalogue to 37 distinct types of myosin — almost double the number known before. Previously unknown myosin tails containing unique combinations of protein domains were revealed, such as the type 3 myosin from the water mould *Phytophthora ramorum*. This has a series of ankyrin (ANK) protein-protein interaction domains followed by a FYVE domain that binds to the phosphoinositide PI(3)P. Moreover, the analysis confirms that no single myosin is common to all organisms, and that the complement of myosins in any given species ranges from two (in *Entamoeba*) to 13 (in *Phytophthora*). The diversity of myosins is likely to be reflected in the range of actin-based movements that a given cell type or organism can generate, and future functional studies of novel myosins may well reveal a wider range of roles for this group of motor proteins than previously suspected.

Is there a single ancestral myosin? The available data can only narrow down the possibilities to the presence of at least three ancestral myosin subfamilies in the eukaryotic cenancestor (Fig. 1). Previous studies had hinted that two of these, the myosin I and MSD myosins, were present in the earliest eukaryotes, and Richards and Cavalier-Smith's more extensive analysis provides a firm basis for this supposition. It also reveals that a third cenancestral myosin group consists of the MyTH4/FERM myosins, which are present not only in amoebae and multicellular animals (metazoans) but also in chromoalveolates.

So how did the different types of myosin evolve? As one might expect, it seems that following the appearance of the major cenancestral groups, the myosins diversified during eukaryotic evolution by gains and losses of protein domains. Notably, class II myosins, some of the best-studied myosins, are not ancient but arose during the evolution of the unikonts (organisms that have a single flagellum), which include the amoeboid, fungal and metazoan lineages. In addition, certain types of myosins were lost in some groups during evolution. For example, myosin I is missing from the plant lineage and, in an extreme example, no myosins could be found in *Trichomonas* and *Giardia* (both of which are primitive unicellular parasites) or red algae. Myosins in these lineages could have either diverged radically from the rest of the family or been lost altogether: given the existence of myosins in other rapidly evolving groups, it seems most likely that they were lost.

Richards and Cavalier-Smith¹ also address the larger question of the nature of the ancestral eukaryote. Their results are consistent with the emerging hypothesis that a fundamental separation between unikonts and bikonts (cells with two flagella), a group that includes plants, chromoalveolates and the excavates, is the earliest evolutionary divergence. They also infer some of the cellular structures that the common ancestor of these two groups must have possessed: it would have had a single cilium, a centriole and a mitochondrion, and would have had the ability to form a pseudopod. This ancestral cell would have had at least three different types of myosin, with the myosin I perhaps regulating formation of the pseudopod to aid cell movement, the MSD myosin contributing to both cell division and organelle movement, and the MyTH4/FERM myosin having a role in adhesion to substrates and perhaps even contributing to the formation of specialized actin-filled

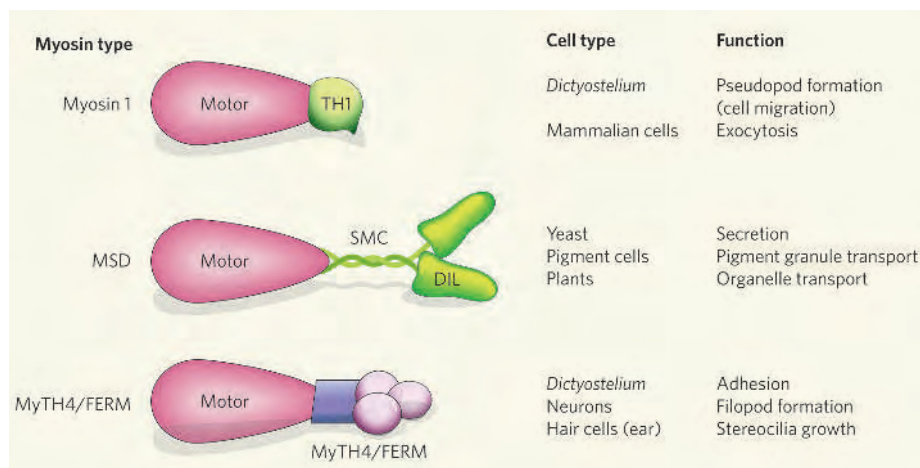


Figure 1 | Three likely ancestral myosins. Richards and Cavalier-Smith¹ propose that there were three ancestral myosins in the earliest eukaryotes, each with distinct tail domain structures. Listed are examples of organisms or cells expressing members of each myosin group and their known functions^{2,3}. The TH1 domain would probably bind to charged lipids and target these myosins to membranes. The SMC domain would promote dimer formation and the DIL and MyTH4/FERM domains would target myosins to their cargo or subcellular location.

projections, such as filopodia, which help the cell explore its environment.

The new myosins uncovered by Richards and Cavalier-Smith will keep motor-protein researchers busy for some time. The first task is to characterize them all functionally, including an analysis of their cellular roles and their motor and structural properties. Comparison of these properties will provide more information about the conservation and diversity of motor function in a range of different cellular contexts. It seems likely that other motor-protein families, such as the kinesins, have a similarly large number of different types, and it will be interesting to see if this is indeed the case. Finally, there is a sixth taxonomic group, the rhizarians, for which sequence data are not yet available, so there could well be more myosins to be discovered there. This taxon includes interesting amoeboid organisms such as the foraminiferans, which are distinguished

from the amoebae by the reticular structure of their pseudopodia. Including analysis of the rhizarians would not only complete the survey of myosin types, but would help to test the current model of eukaryotic evolution. ■

Margaret A. Titus is in the Departments of Genetics, Cell Biology and Development, University of Minnesota, 6-160 Jackson Hall, 321 Church Street SE, Minneapolis, Minnesota 55455, USA.

e-mail: titus004@umn.edu

1. Richards, T. A. & Cavalier-Smith, T. *Nature* **436**, 1113–1118 (2005).
2. Kieck, M. C. & Titus, M. A. in *Molecular Motors* (ed. Schliwa, M.) 3–44 (Wiley VCH, Weinheim, 2003).
3. Krendel, M. & Mooseker, M. S. *Physiology* **20**, 239–251 (2005).
4. Cheney, R. E. *et al.* *Cell Motil. Cytoskel.* **24**, 215–223 (1993).
5. Goodson, H. V. & Spudis, J. A. *Proc. Natl Acad. Sci. USA* **90**, 659–663 (1993).
6. Berg, J. C. *et al.* *Mol. Biol. Cell* **12**, 780–794 (2001).

SYNTHETIC CHEMISTRY

Light on chirality

Yoshihisa Inoue

Reactions that produce only one of two mirror-image forms of a molecule are a hot topic in organic synthesis. A light-driven catalyst provides good results, and the technique could be generally applicable.

Chirality — the non-identity of a molecule with its mirror image — is ubiquitous. It occurs not only in biomolecules (amino acids, sugars, DNA and RNA are examples of chiral molecules), but also in man-made chemicals, materials and drugs. Catalytic asymmetric synthesis — the use of chiral catalysts to transfer and amplify chirality in chemical reactions — has therefore become a central topic in molecular science¹. Bach and colleagues (this issue, page 1139)² now combine two approaches to asymmetric synthesis — the thermal and the photochemical — to control the spatial arrangement of the atoms in a chiral reaction product. The results could be seminal in the field of chiral photochemistry.

In conventional thermal asymmetric synthesis, vibrational energy is supplied to a reaction in the presence of a chiral catalyst or enzyme. This activates ground-state reagent molecules to achieve an asymmetric transformation in which one of two enantiomers — mirror-image forms of a molecule — of a reaction product will be preferentially synthesized. The aim of photochemical asymmetric synthesis is the same, but its tools are different: it uses short-lived, weakly interacting molecular states that have been excited not by heat but by absorbed light. This technique is more difficult to control than its thermal counterpart, and has therefore been less extensively studied, despite its inherent advantages — the low activation

energy required for such reactions and the ability to create unstable molecules unique to photochemical reactions, for example.

Nevertheless, chiral photochemistry, or photochirogenesis, has become an area of rapid growth, particularly in the past 10–15 years^{3–5}. Like other methods in the realm of asymmetric synthesis, photochirogenesis essentially requires a physical or chemical source of chirality that can be transferred to the reaction products. Such sources come in four main varieties. The first is circularly polarized light, which is used in a technique known as absolute asymmetric synthesis — because a product enriched in one enantiomer is formed from a one-to-one mixture of mirror-image precursor molecules without the intervention of a chiral catalyst. This method is not useful for practical synthetic purposes, but has been discussed in relation to a possible extraterrestrial origin of the chiral homogeneity of biomolecules on Earth^{3,6,7}.

The second method, known as the chiral auxiliary strategy, uses a molecular group of a particular chirality that binds covalently to an achiral substrate. The irradiation of this augmented substrate creates a new chiral centre in the substrate, often in such a way that the spatial structure of the new chiral centre in the reaction product is determined by that of the chiral auxiliary. Inevitably, however, this technique requires equal molar quantities of

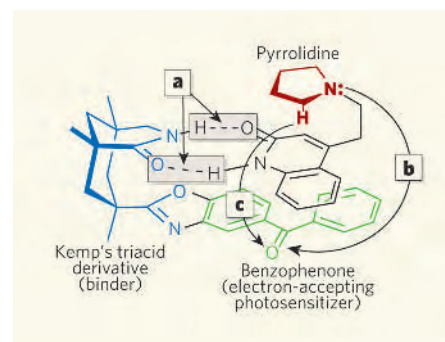


Figure 1 | Choozy catalyst. The photosensitive chiral catalyst molecule developed by Bach *et al.*² latches onto a substrate molecule. **a**, A 'Kemp's triacid derivative' group (blue) that forms part of the catalyst attaches to a selected substrate face using two hydrogen bonds (dotted lines). **b**, When activated by light, a photosensitive benzophenone group (green), also part of the catalyst, accepts an electron from a nitrogen atom in a pyrrolidine group (red) in the substrate, creating a radical-ion pair — a process known as photochemical electron transfer (PET). **c**, This is followed by the transfer of a proton adjacent to the nitrogen, setting in a train a sequence of reactions with an enantioselective result — the preferential formation of one mirror-image version of the reaction.

the chiral source and the substrate.

A third and smarter way to transfer chirality is to use a catalytic chiral complexing agent, which is needed in a smaller molar amount than the substrate. Such an agent binds to the substrate in the ground state to provide a chiral environment for a subsequent photochemical process.

Finally, light-absorbing compounds known as chiral photosensitizers can be used to transfer energy or electrons, along with the chiral information, to a substrate. The advantages of this method are, first, that only a small amount of photosensitizer is needed, and, second, that chirality transfer occurs exclusively in the excited state and is thus unaffected by the binding affinity of the photosensitizer for the substrate in the ground state. This does, however, make controlling the structure of the reaction product more difficult, owing to the weak, short-lived interactions in the excited state.

Bach and colleagues' approach² combines the advantages of the third and fourth methods. The chiral catalyst that they developed (Fig. 1) contains a photosensitizer component — benzophenone — and a group known as a Kemp's triacid derivative, which uses two hydrogen bonds to attach to a specific substrate like a template, favouring one of the two faces of the substrate's molecular plane. In previous studies^{8,9}, the authors had exploited the bulky backbone of a simpler template molecule as a 'picket-fence' to prevent a reagent attacking the substrate from the template side.

In their new catalyst², the benzophenone group accepts an electron from an atom, in

this case nitrogen, of the bound organic substrate (Fig. 1) when activated by light. This creates a radical–ion pair in a process known as photochemical electron transfer, or PET. (PET processes are well known from, among other things, the conversion of solar to chemical energy in plant photosynthesis.) The benzophenone group then forces a molecular group known as a pyrrolidine ring to remain on one side of the substrate, and thus define which side will participate in the further reaction. This process brings about the desired enantioselection, so that one of the possible two mirror images of the reaction product will form preferentially. Bach and colleagues succeed in obtaining a chiral reaction product consisting of up to 70% excess of one enantiomer, with a yield of 52–64% and turnover numbers (a measure of the amount in mole of a product that is obtained with one mole of catalyst) of between 2.1 and 12.2.

The control of chirality is no trivial task, particularly where PET processes are involved. One factor that can significantly disturb the selective formation of one or other enantiomer is the subsequent dissociation, or separation, of the PET-produced radical–ion pair by solvent molecules. This process spoils the chiral recognition between the sensitizer and the substrate; the reaction will eventually create a racemic product — that is, an equal mixture of two enantiomers. It also means that the polar

solvents necessary for PET processes are a mixed blessing: although they accelerate electron transfer and thus increase yield, they simultaneously facilitate dissociation and thus decrease enantioselectivity. A noteworthy previous effort to overcome this trade-off between an excess of one desired enantiomer and high chemical yield was the combined use¹⁰ of a photosensitizer carrying saccharides and a nonpolar solvent. Here, the polar saccharides accelerate PET, whereas the nonpolar solvent prevents the dissociation of the resulting radical–ion pair to secure chiral recognition.

In their current study², Bach and colleagues neatly sidestep the acceleration–dissociation dilemma inherent in PET by using neutral radical species produced in the sequence of electron and proton transfers as intermediates, and a hydrogen bond as a tether, to ensure the stability of the chiral environment (Fig. 1). Their photochirogenic process is composed of four steps: the initial PET to produce a radical–ion pair; the transfer of a proton adjacent to the electron-deficient nitrogen to the benzophenone radical anion; the formation of the resulting radical pair into a hydrocarbon ring (cyclization) within the molecule; and finally, the hydrogen initially transferred to benzophenone is returned to the radical centre of the cyclized product. Throughout the whole process, the dual hydrogen bonds tie

the substrate to the chiral catalyst in close proximity and in the right orientation. Although all the individual techniques were known previously, they have never before been combined to circumvent the acceleration–dissociation problem.

Bach and colleagues thus provide us with a powerful method applicable to PET photochirogenesis. This is certainly a breakthrough in chiral photochemistry — particularly where synthesis is concerned — that will further stimulate research in this rapidly growing area of science. ■

Yoshihisa Inoue is in the Entropy Control Project, ICORP, JST and Department of Applied Chemistry, Osaka University, 2-1 Yamada-oka, Suita 565-0871, Japan.
e-mail: inoue@chem.eng.osaka-u.ac.jp

1. Rouhi, A. M. *Chem. Eng. News* **82**, 47–62 (2004).
2. Bauer, A., Westkämper, F., Grimme, S. & Bach, T. *Nature* **436**, 1139–1140 (2005).
3. Inoue, Y. & Ramamurthy, V. (eds) *Chiral Photochemistry* (Dekker, New York, 2004).
4. Rau, H. *Chem. Rev.* **83**, 535–547 (1983).
5. Inoue, Y. *Chem. Rev.* **92**, 741–770 (1992).
6. Bonner, W. A. *Origins Life Evol. Biosphere* **21**, 59–111 (1991).
7. Bonner, W. A., Greenberg, J. M. & Rubenstein, E. *Origins Life Evol. Biosphere* **29**, 215–219 (1999).
8. Bach, T., Bergmann, H., Grosch, B. & Harms, K. J. *Am. Chem. Soc.* **124**, 7982–7990 (2002).
9. Aechtner, T., Dressel, M. & Bach, T. *Angew. Chem. Int. Edn* **43**, 5849–5851 (2004).
10. Asaoka, S., Wada, T. & Inoue, Y. *J. Am. Chem. Soc.* **125**, 3008–3027 (2003).

BRIEF COMMUNICATIONS

Independent virus development outside a host

Growing two long filamentous tails may help an archaeal virus to survive in a hostile environment.

Viruses are thought to be functionally inactive once they are outside and independent of their host cell¹. Here we describe an exceptional property of a newly discovered virus that infects a hyperthermophilic archaeon growing in acidic hot springs: the lemon-shaped viral particle develops a very long tail at each of its pointed ends after being released from its host cell. The process occurs only at the temperature of the host's habitat (75–90 °C) and it does not require the presence of the host cell, an exogenous energy source or any cofactors. This host-independent morphological development may be a strategy for viral survival in an environment that is unusually harsh and has limited host availability.

The archaeal virus was discovered in an acidic hot spring (85–93 °C; pH 1.5) at Pozzuoli, Italy. In enrichment cultures from the environmental samples, we detected highly unusual, lemon-shaped particles carrying appendages of different lengths at each end (Fig. 1a). The particles were isolated and purified and their viral nature confirmed by replication in cells of the archaeon *Acidianus convivator*; the virus was named *Acidianus* two-tailed virus, or ATV. Surprisingly, particles that emerged from ATV-infected cells growing at 75 °C had no tails (Fig. 1b); no other particles were evident in the infected cell culture during the four days after infection (Fig. 1c). After eight days, however, all visible particles had developed two tails.

Tail-less particles purified from an infected *A. convivator* culture two days after infection were homogeneous in size and shape (Fig. 1c) and remained morphologically unchanged for several months when stored at 4 °C in growth medium or distilled water. But when the particles were incubated at 75 °C in these media in the absence of host cells, a tail developed gradually at each end of the particle (Fig. 1c, d), yielding a population of two-tailed virions that were similar in appearance to the original sample (Fig. 1a) after one week. At 85–90 °C, the optimal temperature for host growth, transformation of the particles was complete within 1 hour. Results from different treatments of tail-less virions confirmed that tail development is an active biological process and is not the result of particle disruption (see supplementary information).

Sequencing and annotation of the circular, 62,730-base-pair, double-stranded DNA genome of ATV revealed a candidate for a tail-

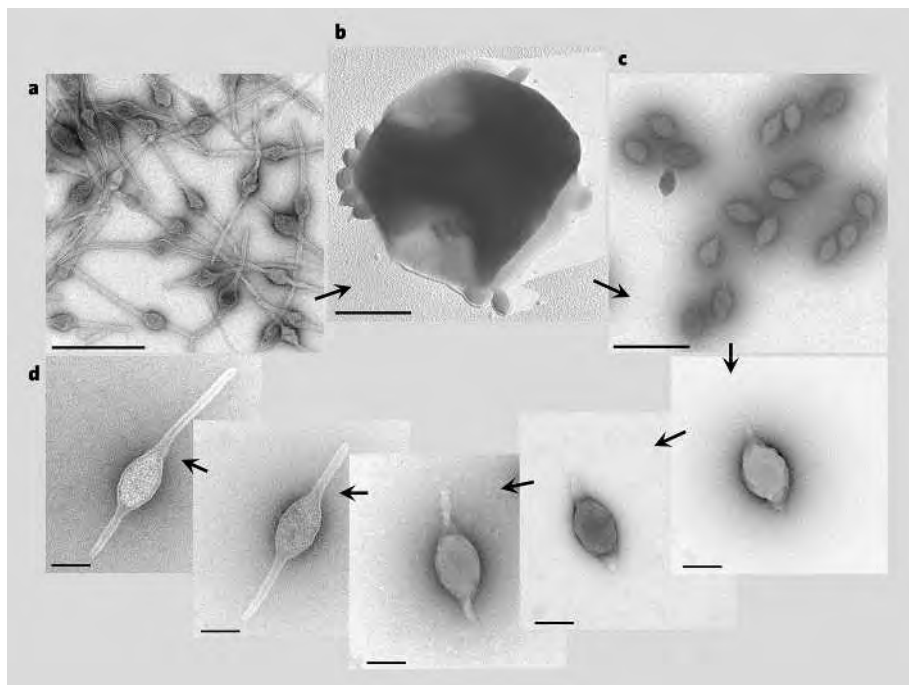


Figure 1 | Electron micrographs of *Acidianus convivator* and different forms of the *Acidianus* two-tailed virus, or ATV. **a, Virions in an enriched sample taken from acidic hot springs in Pozzuoli, Italy (pH 1.5, 85–93 °C). **b**, Extrusion of lemon-shaped virions from an ATV-infected *A. convivator* cell. **c**, Virions in a growing culture of ATV-infected *A. convivator*, 2 days after infection. **d**, Cultured virions after purification (see supplementary information) and incubation at 75 °C for 0, 2, 5, 6 and 7 days (panels from right to left, respectively). All preparations were negatively stained with 3% uranyl acetate, except for **b**, which was platinum-shadowed. Scale bars: **a–c**, 0.5 µm; **d**, 0.1 µm.**

protrusion protein (see supplementary information). This 800-amino-acid protein, one of nine structural proteins identified, has a repeated coiled-coil pattern similar to that of the intermediate-filament proteins^{2,3}. Such proteins participate in the architecture and dynamics of eukaryotic and bacterial cells^{4,5}, and can assemble into filaments in the absence of cofactors or energy sources.

We heterologously expressed the gene encoding this 800-amino-acid protein and found that the purified recombinant protein generated long filamentous structures, 2 nm in width, with features similar to those observed inside the virion tails (results not shown).

There are already several known examples of natural extracellular viral morphogenesis, but these represent either the final steps in particle assembly and maturation, as in retroviruses^{6,7}, or the initial steps of infection, as in tectiviruses⁸, and they are triggered on the host-cell surface concurrently with virus budding or adsorption, respectively. To our

knowledge, ATV is the first example of a virus with host-independent as well as extracellular functional activity.

The development of tails at the temperatures at which hosts are active may constitute part of a strategy for viral survival in unstable and denaturing environmental conditions where host-cell density is low. ATV is the only known virus from hot, acidic habitats that causes lysis of its host cell (see supplementary information). The others maintain a stable relationship with the host cell and its progeny⁹, thereby limiting direct exposure of the virus population to the harsh environment.

This instance of independent morphological development by ATV outside its host cell under exceptional environmental conditions indicates that viruses may be even more biologically sophisticated than previously recognized.

Monika Häring*†, Gisle Vestergaard‡, Reinhard Rachel†, Lanming Chen‡, Roger A. Garrett‡, David Prangishvili*†

*Molecular Biology of the Gene in Extremophiles

Unit, Institut Pasteur, 75724 Paris Cedex 15, France
e-mail: prangish@pasteur.fr

† University of Regensburg, 93053 Regensburg, Germany

‡ Danish Archaea Centre, Institute of Molecular Biology and Physiology, Copenhagen University, 1307 Copenhagen K, Denmark

1. van Regenmortel, M. H. V. in *Seventh Report of the International Committee on Taxonomy of Viruses* (eds Fauquet, C. M. et al.) 3–16 (Academic, San Diego, 2000).
2. Lupas, A., Van Dyke, M. & Stock, J. *Science* **252**, 1162–1164 (1991).
3. Strelkov, S. V., Herrmann, H. & Aebi, U. *BioEssays* **25**, 243–251 (2003).
4. Ausmess, N., Kuhn, J. R. & Jacobs-Wagner, C.

Cell **115**, 705–713 (2003).

5. Hermann, H. & Aebi, U. *Curr. Opin. Cell Biol.* **12**, 79–90 (2000).
6. Pettit, S. C., Everitt, L. E., Choudhury, S., Dunn, B. M. & Kaplan, A. H. *J. Virol.* **78**, 8477–8485 (2004).
7. Swanstrom, R. & Willis, J. W. in *Retroviruses* (eds Coffin, J. M., Hughes, S. H. & Varmus, H. E.) 263–334 (Cold Spring Harbor Lab. Press, New York, 1997).
8. Ackermann, H. W. & Bamford, D. in *Seventh Report of the International Committee on Taxonomy of Viruses* (eds Fauquet, C. M. et al.) 111–116 (Academic, San Diego, 2000).
9. Prangishvili, D., Stedman, K. & Zillig, W. *Trends Microbiol.* **9**, 39–43 (2001).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

doi:10.1038/nature4361101a

GREEN CHEMISTRY

Reversible nonpolar-to-polar solvent

Imagine a smart solvent that can be switched reversibly from a liquid with one set of properties to another that has very different properties, upon command. Here we create such a system, in which a non-ionic liquid (an alcohol and an amine base) converts to an ionic liquid (a salt in liquid form) upon exposure to an atmosphere of carbon dioxide, and then reverts back to its non-ionic form when exposed to nitrogen or argon gas. Such switchable solvents should facilitate organic syntheses and separations by eliminating the need to remove and replace solvents after each reaction step.

Chemical production processes often involve multiple reaction and separation steps, and the type of solvent that is optimum for a particular step may be different from the one needed in the next step. The solvent is therefore usually removed after each step and a new solvent added in preparation for the next, significantly adding to the economic cost and environmental impact of the process. This cumbersome procedure would be unnecessary if a solvent's properties could be adjusted for the following step while still in the reaction vessel, enabling the same solvent to be used for several consecutive reaction or separation steps. Moderate changes in temperature and pressure are incapable of triggering significant changes in the properties of conventional solvents. In contrast, supercritical fluids¹ and CO₂/organic solvent mixtures² can be modified by pressure changes, but unfortunately only above 40 bar.

The reaction we describe reversibly changes the nature and properties of a solvent but occurs under very mild conditions. We reasoned that switching a normal non-ionic liquid to an ionic liquid should induce a change in its properties: ionic liquids are often viscous and always polar, whereas non-ionic solvents are typically non-viscous and vary widely in polarity. We chose CO₂ as the 'switch' to elicit this change because it is a benign agent

and easily removed. (For methods, see supplementary information.)

We found that exposure of a 1:1 mixture of the two non-ionic liquids, namely DBU (1,8-diazabicyclo-[5.4.0]-undec-7-ene) and 1-hexanol, to gaseous CO₂ at one atmosphere and at room temperature causes conversion of the liquid mixture to an ionic liquid (Fig. 1a, b). This is readily converted back into a non-ionic liquid by bubbling N₂ or argon through

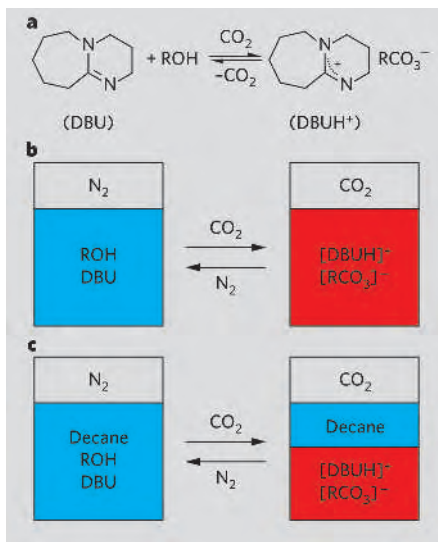


Figure 1 | The 'switching' of a switchable solvent.

a, Protonation of DBU (1,8-diazabicyclo-[5.4.0]-undec-7-ene) in the presence of an alcohol and carbon dioxide is reversed when CO₂ is removed.

b, Polarity switching in the reaction shown in **a**, in which CO₂ causes a nonpolar liquid (shown in blue) mixture of hexanol and DBU to change over one hour into a polar, ionic liquid (shown in red); nitrogen gas reverses the process by stripping out CO₂ from the reaction. **c**, The different polarity of each liquid under the two conditions is illustrated by the miscibility of decane with the hexanol/DBU mixture under nitrogen, before exposure to CO₂; however, decane separates out once the mixture becomes polar in the presence of CO₂. Again, N₂ reverses the process.

the liquid at room temperature or, for a more rapid reaction, at 50 °C. These changes are demonstrated by chemical shifts in key protons, as revealed by ¹H-NMR spectroscopy, and by solvatochromic measurement of the polarity of the solvent before and after exposure to CO₂ (see supplementary information).

The reaction is exothermic and causes a marked increase in the viscosity of the liquid. The choice of alcohol is critical because the 1-hexylcarbonate salt (Fig. 1, right) is a viscous liquid at room temperature, whereas the bicarbonate^{3,4} and methylcarbonate (ref. 5, and A. D. Main, G. E. Fryxell and J. Linehan, unpublished results) salts are solids and so are not candidates for smart solvents.

Our non-ionic liquid is as nonpolar as chloroform, according to measurements using Nile Red as solvatochromic dye (see supplementary information), whereas the liquid under CO₂ is as polar as dimethylformamide or propanoic acid. The polarity changes in this switchable solvent system are demonstrated by testing the solubility of decane, a nonpolar compound, in each liquid: it is miscible with the liquid under N₂ but not with that under CO₂ (Fig. 1c). We conclude that N₂ and CO₂ at 1 bar can be used as triggers of miscibility and immiscibility, respectively.

We have built solvent switchability into molecules that are small enough to be liquid at room temperature. Further examples of switchable solvents, preferably ones less Lewis-basic than DBU, should eventually enable their application in the 'green' production of high-value chemicals such as pharmaceuticals.

Philip G. Jessop*, David J. Heldebrant*, Xiaowang Li*, Charles A. Eckert†, Charles L. Liotta†

*Department of Chemistry, Queen's University, Kingston, Ontario K7L 3N6, Canada

e-mail: jessop@chem.queensu.ca

†Schools of Chemistry and Chemical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0100, USA

1. Jessop, P. G. & Leitner, W. (eds) *Chemical Synthesis using Supercritical Fluids* (VCH/Wiley, Weinheim, 1999).
2. Subramaniam, B. & Busch, D. H. in *Carbon Dioxide Conversion and Utilization* (eds Song, C., Gaffney, A. F. & Fujimoto, K.) 364–386 (ACS, Washington, 2002).
3. Perez, E. R. et al. *J. Org. Chem.* **69**, 8005–8011 (2004).
4. Heldebrant, D. J., Jessop, P. G., Thomas, C. A., Eckert, C. A. & Liotta, C. L. *J. Org. Chem.* **70**, 5335–5338 (2005).
5. Munshi, P., Main, A. D., Linehan, J., Tai, C. C. & Jessop, P. G. *J. Am. Chem. Soc.* **124**, 7963–7971 (2002).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

doi:10.1038/nature4361102a

Corrigendum

Dogs cloned from adult somatic cells

Byeong Chun Lee, Min Kyu Kim, Goo Jang, Hyun Ju Oh, Fibrianto Yuda, Hye Jin Kim, M. Hossein Shamin, Jung Ju Kim, Sung Keun Kang, Gerald Schatten, Woo Suk Hwang *Nature* **436**, 641 (2005)

This communication contains an error in the methods section of the supplementary information. In the description of the fusion protocol on page 3, line 2, electrical pulses were delivered for 15 microseconds, and not for 15 seconds as published.

Reversing histone methylation

Andrew J. Bannister¹ & Tony Kouzarides¹

Histones package DNA, and post-translational modifications of histones can regulate access to DNA. Until recently, histone methylation—unlike all other histone modifications—was considered a permanent mark. The discovery of enzymes that reverse the methylation of lysines and arginines challenges our current thinking on the unique nature of histone methylation, and substantially increases the complexity of histone modification pathways.

In its 'naked' form, DNA is unwieldy and unmanageable for a cell to package. This problem is solved by histones, which compact and control DNA. The many different types of histone modifications (for example, acetylation, methylation, phosphorylation and ubiquitination; reviewed in refs 1–3) regulate DNA-based events in ways that were unimaginable a decade ago. Histone methylation, perhaps more than any other form of modification, has demonstrated the power of modifications over DNA-based functions, regulating fundamental processes such as gene transcription and DNA repair. Furthermore, since the discovery of the first histone methyltransferase⁴, the potential for the methylation 'mark' to control epigenetic events has caught the imagination of workers in this field. However, the recent discovery that methylation can be reversed^{5–7} has shaken the dogma that a 'permanent' methylation mark is necessary for epigenetic control.

Histones may be methylated on either lysine (K) or arginine (R) residues. It is possible that methylation induces alterations in chromatin architecture, either condensing or relaxing its structure. However, a methyl group is relatively small and its addition to lysine or arginine residues does not neutralize their charge, so it is unlikely that methylation alone will significantly affect chromatin structure. It is more likely that it creates binding sites for regulatory proteins that contain specialized binding domains.

Lysine side chains may be mono-, di- or tri-methylated, whereas the arginine side chain may be mono-methylated or (symmetrically or asymmetrically) di-methylated^{3,8}. At present, there are 24 known sites of methylation on histones (17 are lysine residues and 7 are arginine residues). If we take into consideration all three possible methylation states of lysine and arginine, there are potentially 3×10^{11} distinct methylation states of histone proteins. Although all of this combinatorial specificity may not be used, this calculation highlights the vast potential for the regulation of function, and the enormity of the task of understanding how methylation works. Why are there such a huge number of possibilities? Are there specific functions that are controlled by a subset of modifications? Is this combinatorial specificity predictive, like a code? How do specific modifications give rise to appropriate biological outcomes? Here we review what is currently known about methylation and its control of chromatin function, and consider the implications of recent reports indicating that the methylation of histones is a dynamic process.

Methylation of lysines

The most-studied sites of lysine methylation lie in the amino termini of H3 and H4 histone proteins (Table 1). At a first level of characterization, these methylated sites are defined by their presence within a certain type of chromatin, either heterochromatin (a condensed and 'transcriptionally silent' chromatin) or euchromatin (a loosely packed and 'transcriptionally active' chromatin). In certain

cases, the enzymes that mediate the methylation have been shown to direct the formation of specific chromatin states and to be responsible for transcriptional regulation (Table 1).

It is becoming clear that not all heterochromatin is the same with respect to the methylated histones that it contains. The methylated sites on the histones found within heterochromatin (H3K9, H3K27, H3K79 and H4K20) demarcate subdomains; tri-methylated H3K9 and tri-methylated H4K20 are enriched in pericentric heterochromatin, whereas tri-methylated H3K27 is enriched at the inactive X-chromosome^{9–15}. This information could imply the existence of some sort of code, but whether this is predictive, with respect to the chromatin structure formed at these sites, remains to be established.

As with heterochromatin, not all euchromatin is the same. Genes within euchromatin have the potential to be active and are associated with methylated H3K4 and H3K36 histones. When a gene is expressed in yeast, further rounds of histone methylation appear in a localized fashion (enriched at the 5' end of the gene) and in specific forms, primarily tri-methylation (reviewed in ref. 3). A large-scale analysis of human euchromatin indicates that a situation similar to the one in yeast may also occur in mammals¹⁶.

The extent of our knowledge regarding the mechanistic and functional consequences of methylation is limited to the proteins and domains that recognize the modification. Repressive proteins, such as heterochromatin protein 1 (HP1) or the *Drosophila* Polycomb (PC) protein, contain a chromodomain that allows them to specifically recognize the appropriate repressive methylation mark (H3K9 and H3K27 respectively; reviewed in ref. 3), whereas the chromodomain helicase DNA-binding protein 1 (CHD1) activator protein from *Saccharomyces cerevisiae* uses its chromodomain to bind the activating methylated H3K4 (ref. 17). Therefore, the ultimate function of the methyl group is a reflection of the type of protein it has evolved to recruit—either an activator or a repressor of transcription (Fig. 1).

Recently, two domains that are distinct from the chromodomain were shown to bind methylated lysine residues. The Tudor domain within the DNA-repair checkpoint protein p53-binding protein 1 (p53BP1) recognizes methylated H3K79, a widely distributed histone modification in mammalian cells¹⁸. This finding fulfils the prediction for members of the larger Royal Family domain, which were thought to bind methylated lysine¹⁹. The WD40 repeats of the vertebrate transcriptional activator WDR5 also forms a binding site for a methylated lysine, in this case di- and tri-methylated H3K4 (ref. 20). The challenge in the future is to understand how the recruitment of specific proteins to methylated sites mediates the desired biological function.

Almost all methylation marks characterized to date have been shown to have a role in transcription. This monopoly of function is likely to be less a reflection of a unique role for methylation than of a bias in the current research. There is no reason to believe that other DNA functions, such as replication, recombination and repair, are

¹Gurdon Institute and Department of Pathology, Tennis Court Road, Cambridge CB2 1QN, UK.

methylated lysines within histones came about from several lines of research. First, reports from over 30 years ago concluded that methylated lysines have the same half-life as histones. Second, the more recent discovery that methylation at H3K9 is responsible for forming and maintaining heterochromatin (a very stable and heritable chromatin state) bolstered the argument that methylation of histones is a permanent 'epigenetic' mark. Third, the mere fact that a demethylating enzyme had not been discovered, although many workers had searched for it, reinforced the view that methylation was a static process.

This view had a chink in its armour from the very beginning: an enzyme with demethylase activity had been reported by Paik and Kim in 1973 (ref. 28), although this activity was never attributed to a particular protein. Indeed, the reversibility of methylation became apparent a few years ago when antibodies against methylated arginine or methylated lysine residues were used in chromatin immunoprecipitations. These experiments revealed that the methylation of histone residues appeared to be reduced under certain conditions. This prompted the idea that demethylation was a likely possibility, and a proposal was put forward which suggested that such an enzymatic activity would function through an amine oxidase reaction (ref. 8 and Fig. 2).

Recently, the enzyme LSD1 (lysine-specific demethylase 1; also referred to as BHC110 or p110b) was identified, which is able to demethylate a specific lysine (K4) within histone H3 using an amine oxidase reaction⁵. This enzyme had previously been identified in a number of repressor complexes (refs 29, 30 and references therein), a fact that fits well with its ability to demethylate the activating methylation site at H3K4. However, demethylation by LSD1 is limited to mono- or di-methylated H3K4: it cannot demethylate tri-methylated H3K4. This is precisely as predicted for an amine oxidase reaction, yet it is the tri-methylated state that is most associated with active genes. Because the transcription of many genes is dynamic, enzymes capable of removing the tri-methylated state should exist. In addition, enzymes that mediate tri-methylation, such as enhancer of zeste homolog 2 (EZH2), are implicated in cancer, so it is probable that the cell has enzymes to reverse this methylation and counterbalance this potentially dangerous methylation state. Demethylation of tri-methylated lysine would require a distinct set of enzymes to the amine oxidases. Such enzymes will most probably function through a pathway involving a hydroxyl radical attack⁸. As there are no apparent LSD1 homologues in *S. cerevisiae* (ref. 5), it is puzzling how this yeast deals with the high levels of H3K4 methylation it possesses. Is H3K4 methylation irreversible in

S. cerevisiae? Is it reversed by distinct mechanisms? Or are there distantly related LSD1-like demethylases still to be discovered?

The LSD1 demethylase is not part of a big family and does not have many obvious homologues. This is rather surprising as there are many methylated lysines in histones and LSD1 seems to be very specific for H3K4. The answer to this problem may be found in proteins that associate with LSD1. The androgen receptor appears to alter the specificity of LSD1 from H3K4 to H3K9, and thereby converts the demethylase from a repressor to an activator of transcription³¹. Thus, at least for androgen receptor target genes, an H3K9 demethylase has been identified. But what about at other sites of lysine methylation? Are there other LSD1 binding factors that alter the specificity of this demethylase? And what about the demethylation of trimethylated H3K9 in heterochromatin, which is apparently stable? One way to resolve this issue of stability is to evoke a dynamic demethylation of mono- and di-methylated H3K9 (by LSD1-like enzymes) but a relatively 'stable' tri-methylated state. In other words, the enzymes that demethylate tri-methylated H3K9 might be tightly controlled or allosterically inhibited so the modification appears stable.

A final point to make concerning LSD1 is that it has the potential to reverse the methylation of H3K4 performed by the oncogenic mixed-lineage leukaemia 1 (MLL1) methyltransferase³, whose gene is found to be rearranged in leukaemia. Thus, it may be that LSD1 is itself involved in cancer. The logic behind this statement comes from the analogy with the histone acetylation pathway. Acetylases (like p300/CBP) are found to be rearranged in cancer cells, and the enzymes that reverse acetylation (deacetylases) are found to be overexpressed in cancer cells.

Arginine methylation

This modification has been relatively difficult to detect *in vivo*, although the existence of a number of protein arginine methyltransferases (PRMTs) suggest that this is a relatively prevalent modification^{32,33}. Mass spectrometry has shown that arginine methylation is present on purified histones in the mono-methylated state³⁴. *In vitro*, however, enzymes such as coactivator-associated arginine methyltransferase 1 (CARM1; also known as PRMT4) can further catalyse the reaction to a di-methylated form. Whether this di-methylated state is deposited on histones *in vivo* is still unclear. Antibodies have been raised that can recognize di-methylated arginine by chromatin immunoprecipitation but it is now clear that many, if not all, commercially available antibodies cross-react with mono-methylated arginine (our unpublished observations). So it is still unclear whether di-methylation takes place *in vivo*.

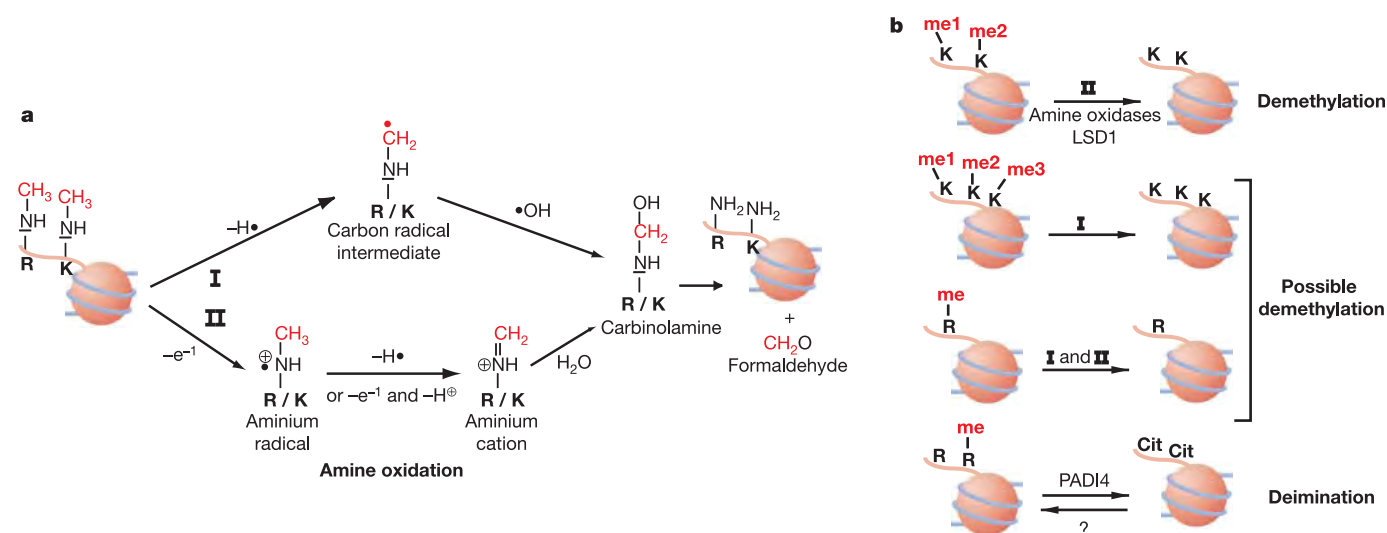


Figure 2 | Reaction mechanisms for methyl group removal. **a**, Two potential chemical reactions (I and II) for the removal of methyl groups (shown in red) from lysine (K) and/or arginine (R) side-chains (adapted from ref. 8). A methylated amine group from the side-chain of each amino acid is shown.

b, Representation of different mechanisms (possible and actual) for removing methyl groups through specific demethylation and deimination processes. I and II relate to the reactions outlined in **a**. me1, me2 and me3 represent mono-, di- and tri-methylated states, respectively; Cit, citrulline.

The methylation of arginine residues has only been linked to active transcription because this modification is only found on chromatin when genes are actively transcribed³³. This modification has been best studied as part of the oestrogen signalling pathway. During transcriptional activation by the oestrogen receptor, arginine methylation of H3 appears transiently and in a cyclical manner³⁵. It is unclear why these cycles of modification take place, but they suggest the existence of enzymes that reverse arginine methylation. It is also unclear how arginine methylation is involved in the activation signal. A methyl-arginine binding protein has yet to be discovered.

Removal of arginine methylation

The search for arginine demethylases over the last few years has been fruitless. However, the fact that lysine demethylases such as LSD1 exist, makes it much more likely that there is an arginine demethylase. The amino oxidase reaction, through which LSD1 works, is predicted to be compatible with the demethylation of methyl-arginines as well as methyl-lysines (Fig. 2). However, an alternative pathway for the reversal of arginine methylation has been proposed⁸ and recently shown to be operational on histones in mammalian cells^{6,7}. This pathway involves the removal of a methyl group from an arginine by the conversion of the methyl-arginine residue into citrulline. This process is termed deimination, since the methyl group is removed along with the imine group of arginine. The enzyme that mediates this reaction, peptidyl arginine deiminase 4 (PADI4), converts unmodified arginine and mono-methylated (but not di-methylated) arginine to citrulline at specific sites on the tail of H3 and H4. This activity of PADI4 is linked to the repression of an oestrogen-controlled gene, *pS2*.

The regulated deposition of citrulline in histones raises a number of issues. First, what is the functional consequence of conversion to citrulline? Is it just a way of removing a methyl group, or does the citrulline itself have a positive role to play in transcription repression? One can imagine proteins being recruited that recognize the citrullinated histones specifically, or perhaps the conformation of the histone being altered. Second, citrulline deposition appears to be transient during gene expression^{6,7}, so how does citrulline get converted back to arginine or methyl-arginine? Is there a rapid replacement of histones by unmodified variants, or are there specific enzymes that mediate the reverse reaction? Enzymes that convert non-peptidyl citrulline to arginine are known to exist, so peptidyl amino transferases may function on citrullinated histones.

Discussion

Deimination and demethylation are both processes that reverse methylation but they are unlikely to be redundant. Even though methylation of arginines and lysines is evolutionarily conserved from mammals to yeast, deiminating enzymes appear to be restricted to higher organisms. The tissue-specific expression pattern of deiminases, and the connection between citrulline and human disease³⁶, points to a specialized role for deimination in controlling developmental processes. Perhaps the post-translational deposition of a non-coded residue (citrulline) in place of a modified residue (methyl-arginine) may be a process that has evolved to provide an additional level of control to a complex organism. If this is so, then post-translational conversion of amino acids other than arginine may well take place. In contrast to deimination, demethylation is likely to be an activity that is needed for all organisms in which chromatin modifications are found. The identification of many new demethylases is on the horizon, given the plethora of different methylation sites. Indeed, since arginine methylation occurs in lower organisms, but deimination does not, the chances are also high that arginine demethylases will also be discovered. One thing is for certain—if there is a barrel of enzymes that modify histones, we have not yet reached the bottom.

- Peterson, C. L. & Laniel, M. A. Histones and histone modifications. *Curr. Biol.* **14**, R546–R551 (2004).
- Margueron, R., Trojer, P. & Reinberg, D. The key to development: interpreting the histone code? *Curr. Opin. Genet. Dev.* **15**, 163–176 (2005).
- Rea, S. *et al.* Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**, 593–599 (2000).
- Shi, Y. *et al.* Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**, 941–953 (2004).
- Cuthbert, G. L. *et al.* Histone deimination antagonizes arginine methylation. *Cell* **118**, 545–553 (2004).
- Wang, Y. *et al.* Human PAD4 regulates histone arginine methylation levels via demethyliminol. *Science* **306**, 279–283 (2004).
- Bannister, A. J., Schneider, R. & Kouzarides, T. Histone methylation: dynamic or static? *Cell* **109**, 801–806 (2002).
- Martens, J. *et al.* The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *EMBO J.* **24**, 800–812 (2005).
- Peters, A. H. *et al.* Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. *Mol. Cell* **12**, 1577–1589 (2003).
- Schotta, G. *et al.* A silencing pathway to induce H3–K9 and H4–K20 trimethylation at constitutive heterochromatin. *Genes Dev.* **18**, 1251–1262 (2004).
- Plath, K. *et al.* Role of histone H3 lysine 27 methylation in X inactivation. *Science* **300**, 131–135 (2003).
- Kohlmaier, A. *et al.* A chromosomal memory triggered by Xist regulates histone methylation in X inactivation. *PLoS Biol.* **2**, E171 (2004).
- Rougeulle, C. *et al.* Differential histone H3 Lys-9 and Lys-27 methylation profiles on the X chromosome. *Mol. Cell Biol.* **24**, 5475–5478 (2004).
- Rice, J. C. *et al.* Histone methyltransferases direct different degrees of methylation to define distinct chromatin domains. *Mol. Cell* **12**, 1591–1598 (2003).
- Bernstein, B. E. *et al.* Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120**, 169–181 (2005).
- Pray-Grant, M. G., Daniel, J. A., Schieltz, D., Yates, J. R. III & Grant, P. A. Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation. *Nature* **433**, 434–438 (2005).
- Huyen, Y. *et al.* Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks. *Nature* **432**, 406–411 (2004).
- Maurer-Stroh, S. *et al.* The Tudor domain 'Royal Family': Tudor, plant Agnet, Chromo, PWWP and MBT domains. *Trends Biochem. Sci.* **28**, 69–74 (2003).
- Wysocka, J. *et al.* WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell* **121**, 859–872 (2005).
- Sanders, S. L. *et al.* Methylation of histone H4 lysine 20 controls recruitment of Crb2 to sites of DNA damage. *Cell* **119**, 603–614 (2004).
- Freitag, M. & Selker, E. U. Controlling DNA methylation: many roads to one modification. *Curr. Opin. Genet. Dev.* **15**, 191–199 (2005).
- Fraga, M. F. *et al.* Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. *Nat. Genet.* **37**, 391–400 (2005).
- Pal-Bhadra, M. *et al.* Heterochromatic silencing and HP1 localization in *Drosophila* are dependent on the RNAi machinery. *Science* **303**, 669–672 (2004).
- White, S. A. & Allshire, R. C. Loss of Dicer fowls up centromeres. *Nature Cell Biol.* **6**, 696–697 (2004).
- Elgin, S. C. & Grewal, S. I. Heterochromatin: silence is golden. *Curr. Biol.* **13**, R895–R898 (2003).
- Verdel, A. *et al.* RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* **303**, 672–676 (2004).
- Paik, W. K. & Kim, S. Enzymatic demethylation of calf thymus histones. *Biochem. Biophys. Res. Commun.* **51**, 781–788 (1973).
- Hakimi, M. A. *et al.* A core-BRAF35 complex containing histone deacetylase mediates repression of neuronal-specific genes. *Proc. Natl Acad. Sci. USA* **99**, 7420–7425 (2002).
- Shi, Y. J. *et al.* Coordinated histone modifications mediated by a CtBP co-repressor complex. *Nature* **422**, 735–738 (2003).
- Metzger, E. *et al.* LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. *Nature advance online publication*, 3 August 2005 (doi:10.1038/nature04020).
- Boisvert, F. M., Chenard, C. A. & Richard, S. Protein interfaces in signalling regulated by arginine methylation. *Sci. STKE* doi:10.1126/stke.2712005re2 (2005).
- Lee, D. Y., Teyssier, C., Strahl, B. D. & Stallcup, M. R. Role of protein methylation in regulation of transcription. *Endocr. Rev.* **26**, 147–170 (2005).
- Strahl, B. D. *et al.* Methylation of histone H4 at arginine 3 occurs *in vivo* and is mediated by the nuclear receptor coactivator PRMT1. *Curr. Biol.* **11**, 996–1000 (2001).
- Metivier, R. *et al.* Estrogen receptor- α directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. *Cell* **115**, 751–763 (2003).
- Vossenaar, E. R., Zendman, A. J., van Venrooij, W. J. & Pruijn, G. J. PAD, a growing family of citrullinating enzymes: genes, features and involvement in disease. *Bioessays* **25**, 1106–1118 (2003).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.K. (tk106@mole.bio.cam.ac.uk).

- Fischle, W., Wang, Y. & Allis, C. D. Histone and chromatin cross-talk. *Curr. Opin. Cell Biol.* **15**, 172–183 (2003).

Evolution of helium isotopes in the Earth's mantle

Cornelia Class¹ & Steven L. Goldstein^{1,2}

Degassing of the Earth's mantle through magmatism results in the irreversible loss of helium to space, and high $^3\text{He}/^4\text{He}$ ratios observed in oceanic basalts have been considered the main evidence for a 'primordial' undegassed deep mantle reservoir. Here we present a new global data compilation of ocean island basalts, representing upwelling 'plumes' from the deep mantle, and show that island groups with the highest primordial signal (high $^3\text{He}/^4\text{He}$ ratios) have striking chemical and isotopic similarities to mid-ocean-ridge basalts. We interpret this as indicating a common history of mantle trace element depletion through magmatism. The high $^3\text{He}/^4\text{He}$ in plumes may thus reflect incomplete degassing of the deep Earth during continent and ocean crust formation. We infer that differences between plumes and the upper-mantle source of ocean-ridge basalts reflect isolation of plume sources from the convecting mantle for $\sim 1\text{--}2$ Gyr. An undegassed, primordial reservoir in the mantle would therefore not be required, thus reconciling a long-standing contradiction in mantle dynamics.

The presence of primordial ^3He in mantle gases has been considered the primary evidence for the contribution of undegassed, primordial lower mantle to upwelling plumes that generate ocean island basalts (OIB)^{1–4}, and forms a strong basis for models of two-layer mantle convection^{2,5,6}. On the basis of their helium isotope ratios relative to mid-ocean-ridge basalts (MORB, with $^3\text{He}/^4\text{He} = 8 \pm 1R_A$, where R_A is the atmospheric ratio of 1.4×10^{-6})⁷, so-called 'high- $^3\text{He}/^4\text{He}$ ' and 'low- $^3\text{He}/^4\text{He}$ ' mantle plumes have been distinguished². It has been argued that low- $^3\text{He}/^4\text{He}$ plumes reflect degassed, recycled material from the surface of the Earth dominated by radiogenic ^4He , whereas high- $^3\text{He}/^4\text{He}$ plumes reflect contributions from undegassed mantle^{2,4,8–12}. However, it has been long recognized from trace element and Sr–Nd–Pb–Hf isotope ratios in high- $^3\text{He}/^4\text{He}$ OIB that mantle plume sources do not represent undegassed primitive material^{4,13–16}. In addition, geophysical evidence for whole-mantle convection^{17–20} makes the isolation of a primordial, undegassed high- $^3\text{He}/^4\text{He}$ component in the lower mantle problematic. Thus, the preservation of significant primordial ^3He in OIB remains difficult to explain.

Here we present a global compilation of OIB and show that primordial ^3He in plume sources is represented by a common component, which shows striking chemical and Pb–Sr–Nd isotopic similarities to MORB mantle. The data indicate that high- $^3\text{He}/^4\text{He}$ plume sources, like the upper-mantle source of MORB, chemically complement the continental crust and have been depleted in incompatible elements through Earth history through melt extraction. The observations are explained by means of a model of mantle helium evolution that does not require the existence of a primordial, undegassed high- $^3\text{He}/^4\text{He}$ reservoir.

Helium isotope systematics of oceanic basalts

Our new data compilation of Th, U and La abundances and Sr–Nd–Pb isotope ratios of OIB and MORB are based on global data from the online GEOROC (<http://georoc.mpch-mainz.gwdg.de/>) and PetDB (www.petdb.org) databases, and the USGS noble gas database²¹. To obtain a data set representative of the upper-mantle sources of MORB, only analyses on fresh glasses from a water

depth of greater than 2000 m are used, thus filtering out plume influenced samples. To obtain a representative data set of mantle plume sources, only basaltic rocks from the constructional phase of ocean islands are used (details in Supplementary Information). Basalts from individual islands often show a range in $^3\text{He}/^4\text{He}$ ratios, which often does not covary with Sr–Nd–Pb isotope ratios. This 'decoupling' is not well understood, and it has been suggested that helium as a noble gas behaves differently from lithophile trace elements within the deep mantle or during partial melting processes, magma transport and storage^{14,22}. In addition, Hawaii, Iceland and Galapagos basalts show a range in $^3\text{He}/^4\text{He}$ ratios, from high values to MORB-like values, which partly seem to reflect entrainment of ambient MORB-type mantle, but also require heterogeneity in the plumes (recently reviewed for Hawaii in ref. 23). This contribution uses the highest helium isotope ratios in individual oceanic islands ($^3\text{He}/^4\text{He}_{\text{max}}$) to constrain the nature of the high- $^3\text{He}/^4\text{He}$ component in the associated mantle plumes. We divide the islands into four groups on the basis of the highest $^3\text{He}/^4\text{He}$ ratios from mineral separates and glass: first, 'low $^3\text{He}/^4\text{He}$ ' ($^3\text{He}/^4\text{He} < 7R_A$); second, 'MORB-like $^3\text{He}/^4\text{He}$ ' ($8 \pm 1R_A$); third, 'moderately high $^3\text{He}/^4\text{He}$ ' ($9\text{--}15R_A$); and fourth, 'high $^3\text{He}/^4\text{He}$ ' ($^3\text{He}/^4\text{He} > 15R_A$). These designations enable us to evaluate ocean islands and their plume sources on the basis of their $^3\text{He}/^4\text{He}$ ratios, despite the paucity of combined He–Sr–Nd–Pb–Hf isotope data on individual samples.

The four OIB groups show strong relationships to their Pb–Nd isotope ratios (Fig. 1). Lower $^{143}\text{Nd}/^{144}\text{Nd}$ ratios are associated with lower $^3\text{He}/^4\text{He}$ (Fig. 1). For a constant $^{143}\text{Nd}/^{144}\text{Nd}$ ratio, higher $^{206}\text{Pb}/^{204}\text{Pb}$ ratios are associated with groups with lower $^3\text{He}/^4\text{He}$. Thus, once Nd isotopes are 'normalized' out, high $^3\text{He}/^4\text{He}$ ratios in ocean islands are associated with low $^{206}\text{Pb}/^{204}\text{Pb}$ ratios (Fig. 1, dark blue symbols). The only known exception, Samoa, is a 'high $^3\text{He}/^4\text{He}$ ' plume with low $^{143}\text{Nd}/^{144}\text{Nd}$ ratios and is discussed below. Because ^{206}Pb , ^{207}Pb , ^{208}Pb and ^4He are generated by the radioactive decay of U and Th, the global systematics indicate a direct relationship between $^3\text{He}/^4\text{He}$ and either the plume source formation age or their Th and U content (with older plume sources

¹Lamont-Doherty Earth Observatory of Columbia University, ²Department of Earth and Environmental Sciences, Columbia University, Palisades, New York 10964, USA.

or those with higher Th + U having lower $^3\text{He}/^4\text{He}$ ratios). Here we constrain the Th and U contents of plume sources.

Thorium and lanthanum are both highly incompatible elements and therefore Th/La ratios of basaltic rocks are expected to closely approximate the melt sources. High Th/La therefore indicates Th

enrichment relative to La in the mantle sources. With a few exceptions, mostly from the Society Islands (Fig. 2a, plus signs), Th/La and $^{208}\text{Pb}/^{204}\text{Pb}$ covary positively in OIB, which is consistent with radiogenic ingrowth of ^{208}Pb due to Th decay. The same is true for U/La (not shown), but the data are more scattered owing to the susceptibility of U to alteration by weathering. The Th contents in OIB vary by more than 10^2 and are positively correlated with Th/La (Fig. 2b). Whereas the high-Th end of the OIB array is dominated by relatively

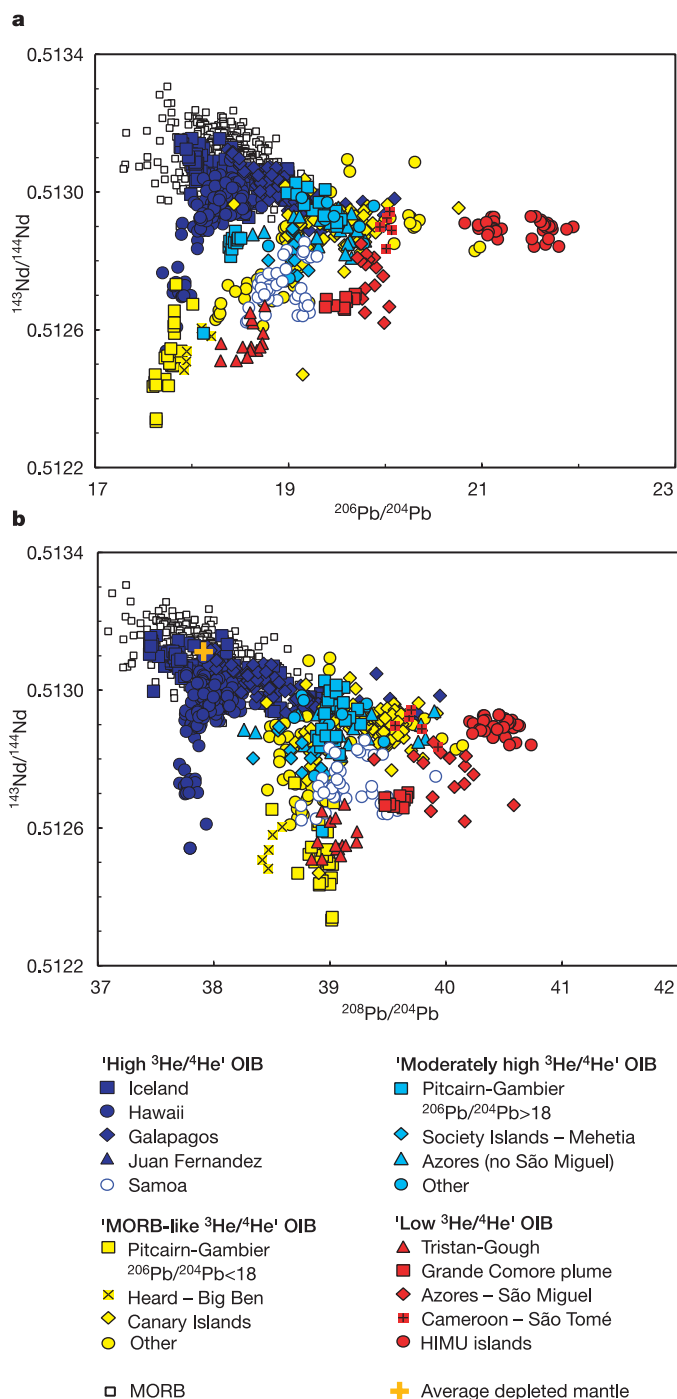


Figure 1 | Nd-Pb isotopes of OIB and MORB. High $^3\text{He}/^4\text{He}$ OIB have the most MORB-like Nd and the lowest Pb isotope ratios. Plots of $^{206}\text{Pb}/^{204}\text{Pb}$ (a) and $^{208}\text{Pb}/^{204}\text{Pb}$ (b) against $^{143}\text{Nd}/^{144}\text{Nd}$ are shown. OIB and MORB data are from the GEOROC and PetDB databases⁴⁹, respectively. Symbol colours correspond to the helium isotope groups of global plumes (see the text and Supplementary Information). Orange plus sign, estimate of average depleted (MORB) mantle⁴⁴. Other 'moderately high $^3\text{He}/^4\text{He}$ ' OIB include Cape Verde northern Islands, Réunion. Other 'MORB-like $^3\text{He}/^4\text{He}$ ' OIB include Austral-Cook Islands Rarotonga and Rurutu younger series, Cape Verde southern Islands, Society seamounts Tehetia, Rocard.

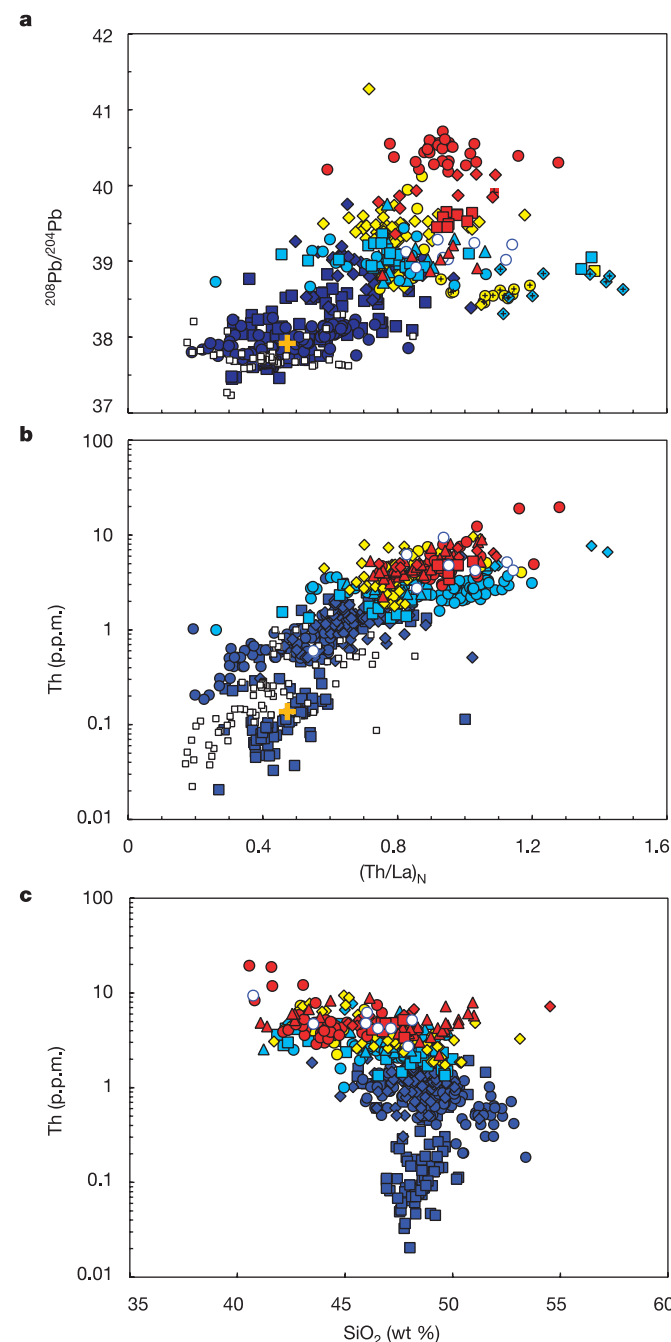


Figure 2 | Th, Th/La, SiO_2 and helium isotope ratios of OIB and MORB. High $^3\text{He}/^4\text{He}$ OIB have MORB-like Th contents and $(\text{Th}/\text{La})_N$ ratios (normalized to primitive mantle⁴⁵). **a, b**, Plots of $^{208}\text{Pb}/^{204}\text{Pb}$ (a) and Th content (b) against $(\text{Th}/\text{La})_N$ in OIB and MORB. **c**, Plot of Th against SiO_2 in OIB. For Th, only samples with $5\text{wt}\% < \text{MgO} < 15\text{wt}\%$ are shown. Symbols and data sources are as in Fig. 1. Society Island samples fall off the main trend in **a** (black plus signs) but not in **b** and **c**. 'Average depleted MORB' (orange plus sign) is estimated to be tenfold the depleted MORB mantle composition⁴⁴ for Th.

low-degree alkali basaltic partial melts, and the low-Th end by high-degree tholeiitic partial melts, the huge range in Th content cannot be solely the result of variable degrees of melting. For example, if global plume sources had similar Th abundances, and tholeiites with 0.5 p.p.m. Th were formed by 6% partial melting, then alkali basalts with about 10 p.p.m. Th would be formed by only about 0.3% partial melting, which is unreasonably low for such rocks. Moreover, lower-degree melts generally have lower silica contents than higher-degree melts, and the absence of a correlation between the SiO_2 and Th content among the helium groups (Fig. 2c) shows that the global range of Th contents in OIB is not the result of different degrees of partial melting. Rather, the large range reflects variable enrichment of Th in the mantle sources of plumes.

A compilation of samples containing both $^3\text{He}/^4\text{He}$ and Th data clearly confirms that plumes with high $^3\text{He}/^4\text{He}_{\text{max}}$ erupt lavas with lower Th contents (Fig. 3a). However, this figure shows that only a small fraction of samples have both $^3\text{He}/^4\text{He}$ and Th analyses. There are no Th data on the samples with the highest $^3\text{He}/^4\text{He}$ ratios from

the 'high $^3\text{He}/^4\text{He}$ ' group, and the 'MORB-like $^3\text{He}/^4\text{He}$ ' group is not represented at all (the range of $^3\text{He}/^4\text{He}$ for each island is illustrated by bars). In Fig. 3b, all measured $^3\text{He}/^4\text{He}$ ratios are plotted against average Th contents of individual islands or plumes. With the exception of Samoa (blue open circles) and Mehetia in the Society Islands (light blue diamonds), $^3\text{He}/^4\text{He}_{\text{max}}$ ratios from individual islands show an inverse relation with Th abundance (and U; compare ref. 16).

Samoa is an outlier to these global systematics, erupting lavas with high Pb and He isotope ratios, and high Th and U. Samoa is also an exceptional plume in its unusual tectonic setting at the northern terminus of the Tonga subduction zone. Farley²⁴ noted the unusual combination of high $^3\text{He}/^4\text{He}$ and Pb-Sr-Nd isotope ratios indicating long-term incompatible element enrichment, and suggested that sediment from the subduction zone became rapidly cycled through the mantle, produced only a small amount of radiogenic ^4He over this short timescale, and became mixed into the rising 'high $^3\text{He}/^4\text{He}$ ' Samoa plume. Because Samoa is such a clear exception to the general rule, we exclude it from the discussion below.

Recycled components and $^3\text{He}/^4\text{He}$ of plumes

There is ample evidence that plume sources contain recycled oceanic crust and sediment that affect their Sr-Nd-Pb-Hf isotope ratios and contribute trace elements such as Th and $\text{U}^{15,25}$. Because helium is degassed during ocean crust formation, alteration and subduction²⁶, the recycled component cannot provide significant primordial ^3He , but it would contain radiogenic ^4He produced after degassing by Th + U decay⁷. The inverse global relationship between $^3\text{He}/^4\text{He}_{\text{max}}$ and Th abundances (Fig. 3) indicates that plumes may contain both a primordial helium component and extra ^4He from Th + U decay in recycled oceanic crust. This is illustrated by the curves in Fig. 3b, where the $^3\text{He}/^4\text{He}$ ratios reflect combinations of radiogenic ^4He from old (1.5 or 0.8 Gyr old) recycled oceanic crust with variable Th plus primordial helium from small amounts (1% or 2%) of 'bulk silicate Earth'. The curves fit the inverse relationship between $^3\text{He}/^4\text{He}_{\text{max}}$ and Th abundances for all helium groups (with Samoa as an exception), which requires similar ^3He abundances in the plume sources (within a factor of about 2). The Th abundances of the 'MORB-like $^3\text{He}/^4\text{He}$ ' and 'low $^3\text{He}/^4\text{He}$ ' groups overlap (Figs 2b and 3b), and this might be explained by differences in the age of formation of the plume sources.

Thus, although some of the variability probably reflects differences in plume source formation age, the global variation of $^3\text{He}/^4\text{He}_{\text{max}}$ and Th can be explained by a small but similar amount of primordial helium plus radiogenic production of ^4He as a function of variable Th + U in global plume sources. Neon data are consistent with this conclusion (see Supplementary Information).

Mantle depletion and high $^3\text{He}/^4\text{He}$ in plumes

The OIB-MORB data compilation shows that the high- $^3\text{He}/^4\text{He}$ component of mantle plumes is a global endmember with Th and U abundances and Sr-Nd isotope ratios similar to MORB (Figs 1–3), indicating that the high- $^3\text{He}/^4\text{He}$ endmember is 'depleted mantle'. This observation is strongly supported by the depleted-mantle composition of the highest $^3\text{He}/^4\text{He}$ endmember for mantle plumes as measured in Baffin Island basalts, associated with the early Iceland plume¹². These have $^3\text{He}/^4\text{He}$ ratios up to $50R_A$ and Sr-Nd isotope ratios and La/Sm in the range of those of depleted MORB. The Baffin data are not shown in Figs 1 and 2 because trace-element abundances and Pb isotope ratios have not yet been published¹². The 'depleted mantle' character of the high- $^3\text{He}/^4\text{He}$ endmember in OIB is also consistent with previous interpretations of Sr-Nd-Hf isotope arrays from individual ocean islands, which seem to converge at isotope ratios reflecting long-term depletion of lithophile elements, which has been termed the 'focus zone' or FOZO³ and is similar in concept and composition to 'primitive He mantle' or PHEM⁴. FOZO also has a low $^{187}\text{Os}/^{188}\text{Os}$, indicative of peridotitic mantle²⁷. The global data

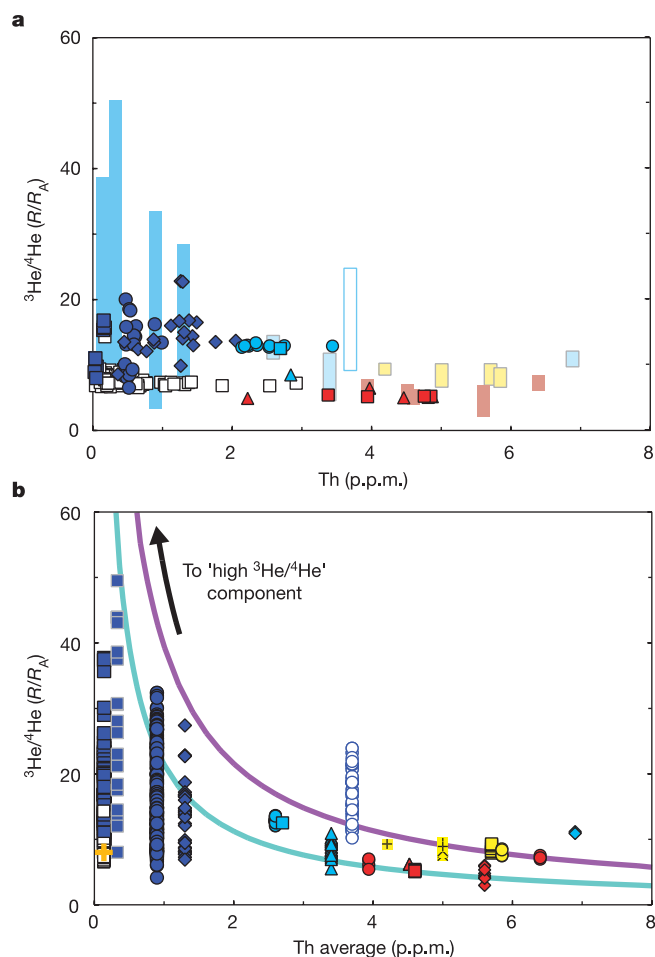


Figure 3 $^3\text{He}/^4\text{He}$ - Th of OIB and MORB. **a**, Symbols show available combined Th- $^3\text{He}/^4\text{He}$ data. Bars show data range from **b**. **b**, Plot of measured $^3\text{He}/^4\text{He}$ against average Th of different plumes and MORB. $^3\text{He}/^4\text{He}_{\text{max}}$ reflect ^4He production from Th + U in recycled components plus a uniform amount of high- $^3\text{He}/^4\text{He}$ component as shown by curves with variable plume source ages plus primitive mantle contribution: blue, 1.5 Gyr (1%); purple, 1.5 Gyr (2%) or 0.8 Gyr (1%). OIBs are assumed to reflect 6% melts with average Th/U = 3.45; primitive mantle $^3\text{He}/^4\text{He}_{4.55 \text{ Gyr}} = 230R_A$; $^4\text{He}_{4.55 \text{ Gyr}} = 1.2 \text{ nmol g}^{-1}$; Th and U from ref. 45. Data sources, filters and symbols as in Figs 1 and 2 plus additional data from the USGS noble gas database²¹; blue squares with grey outline, Baffin Island (early Iceland plume); yellow squares containing plus sign, Rurutu Recent Series and Rarotonga.

are fully consistent with FOZO's representing depleted MORB-like peridotitic mantle with some recycled oceanic crust.

Whereas the upper-mantle source of MORB has lower $^3\text{He}/^4\text{He}$ ratios than the high- $^3\text{He}/^4\text{He}$ endmember of mantle plumes, the MORB $^3\text{He}/^4\text{He}$ value of $8 \pm 1R_A$ (ref. 7), shows that the MORB mantle still contains a significant component of primordial ^3He . Most attempts to explain $^3\text{He}/^4\text{He}$ in MORB have argued for an addition of ^3He to the degassed MORB mantle through a plume flux^{6,28,29} or an independent flux of primordial helium from the lower mantle²². We suggest that the striking chemical and isotopic similarities between depleted MORB mantle and the high- $^3\text{He}/^4\text{He}$ endmember in OIB strongly indicates that their formation is directly related.

A role for high $^3\text{He}/(\text{U} + \text{Th})$ OIB reservoirs has previously been suggested^{12,16,30,31}. Anderson³⁰ assumes that $^3\text{He}/^4\text{He}$ does not covary with other chemical tracers, whereas Figs 1 and 2 show that strong systematic relationships do indeed exist. It has been suggested¹² that depleted upper mantle is re-gassed near the boundary between upper and lower mantle. We suggest that this is problematic because some plumes appear to originate from the core/mantle boundary³² and because plumes have similar concentrations of primordial ^3He , as shown above. The small range of ^3He abundances globally also poses problems for models of recycling of oceanic lithosphere stripped of rare gases into the lower mantle^{16,31}. A well-mixed mantle 'plum pudding' based on two-stage melting of primitive mantle³³ could result in a small range of ^3He abundances, but if this model were correct the high- $^3\text{He}/^4\text{He}$ endmember of OIB should trend towards primitive rather than depleted-mantle compositions (Figs 1 and 2). In addition, all these models^{16,31,34} require primordial, undegassed regions within the lower mantle to preserve primordial ^3He . The 'missing Ar' argument³⁵ concludes that a gas-rich reservoir in the Earth's mantle represents about 50% of its mass and has been used as strong support for a layered mantle, but this is obviated if recent estimates of lower K abundances for the silicate Earth are correct^{30,36,37}. Given the chemical and isotopic similarities between MORB and the high- $^3\text{He}/^4\text{He}$ endmember in OIB, and geophysical evidence for whole-mantle convection^{17–20}, we suggest that high $^3\text{He}/^4\text{He}$ in plumes can be explained by whole-mantle convection and the formation of continents and oceanic crust.

Recent work on noble-gas partition coefficients indicate that they are not extracted with nearly 100% efficiency but instead behave like highly incompatible elements during melting, with a small fraction remaining in melt residues^{38,39}. This allows for some primordial ^3He to be retained in mantle volumes that have been depleted by melt events such as the formation of oceanic crust. We propose that the similarity between MORB and the high- $^3\text{He}/^4\text{He}$ OIB in terms of Th and U contents, Th/La ratios and Sr-Nd-Pb isotope ratios (Figs 1–3), as well as the elevated $^3\text{He}/^4\text{He}$ of MORB (even if lower than 'high $^3\text{He}/^4\text{He}$ ' OIB), could be related through incomplete degassing during ocean and continental crust formation by a mantle undergoing whole-mantle convection. We present a model illustrating how this might work. An initial degassing stage during accretion of the Earth is constrained by xenon isotopes⁴⁰; we consider it to be the event that set up the initial conditions. In this model, the mantle is depleted in incompatible elements (including Th and U) by continent formation over an interval of 4.4–2.7 Gyr, and $^3\text{He}/^4\text{He}$ ratios in this 'depleted mantle' evolve through the decay of Th and U and helium degassing associated with ocean crust formation. The mean age of present-day continental crust is probably closer to 2 Gyr (see ref. 41, for example) but this younger age partly reflects recycling of continental crust to the mantle through sediment subduction and tectonic erosion of continental margins. The higher $^3\text{He}/^4\text{He}$ ratios in plumes than in MORB reflect isolation from the convecting mantle for significant geological intervals. The isolation allows plume sources to retain more ^3He relative to the upper-mantle MORB source, which continually undergoes degassing as it forms oceanic crust. Thus, in this model both MORB and OIB sources have been

significantly degassed through magmatism due to ocean crust and continent formation, but the high- $^3\text{He}/^4\text{He}$ endmember in mantle plumes represents depleted mantle that was 'less degassed' through isolation most probably somewhere in the lower mantle.

The helium evolution model is illustrated in Fig. 4 (calculations and model parameters in Supplementary Methods). The $^3\text{He}/^4\text{He}$ ratio of primitive mantle decreases owing to radioactive production of ^4He (dark blue curve), which would reach about $95R_A$ were it not for the formation of continents and ocean crust. Degassing rates through geological time are under-constrained (details in Supplementary Methods). The most solid constraint is the degassing rate of the oceanic mantle today⁴², which in our model needs to be adjusted with more vigorous degassing in the past. In the example shown, the degassing rate, which reflects both the helium abundance and the rate of ocean crust formation, was 34-fold higher 4.4 Gyr ago. A curve illustrating the effects of degassing on the upper-mantle source of MORB is shown in Fig. 4 (red curve). The evolution of a 'high- $^3\text{He}/^4\text{He}$ component' in plumes is also shown (yellow curve), which is derived from the 'depleted' mantle but becomes isolated from degassing in a boundary layer. In this example the isolation occurred about 1.5 Gyr ago. In contrast to the continuously degassing upper mantle, the $^3\text{He}/^4\text{He}$ ratio of this 'high- $^3\text{He}/^4\text{He}$ component' declines much less and maintains a high present-day value of about $72R_A$.

Figure 4 thus shows that in general terms the high $^3\text{He}/^4\text{He}$ ratios in OIB can be explained through normal mantle differentiation processes. Continent formation results in the generation of a mantle depleted in incompatible elements. The mantle evolves to lower $^3\text{He}/^4\text{He}$ values (as observed in MORB today) through melting and degassing at ocean ridges, re-mixing of degassed mantle by convection and continuous formation of ^4He by radioactive decay. If a portion of this mantle is isolated from degassing, the $^3\text{He}/^4\text{He}$ ratio will remain high, as observed in plume sources, with a value between those of primitive mantle and MORB.

On the basis of this general degassing model, ages of plume sources can be made a dependent variable and calculated (Fig. 5). We calculate Th and U contents of OIB sources assuming that lavas reflect 6% partial melting of peridotite. Plume sources contain variable amounts of recycled components of variable compositions¹⁵.

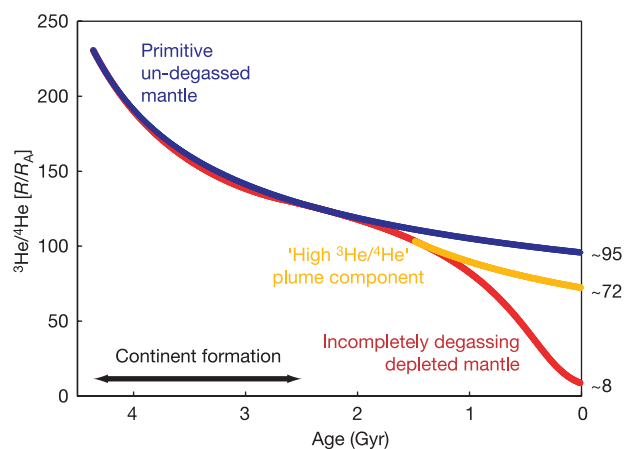


Figure 4 | Helium isotope evolution of mantle reservoirs. Illustration of the incomplete degassing model. Blue line, primitive, undegassed Earth's mantle: $^3\text{He}/^4\text{He} = 230R_A$ (ref. 46), $[^3\text{He}] = 1.5 \times 10^{11} \text{ atoms g}^{-1}$ based on refs 47, 48. Th and U abundances: primitive mantle⁴⁵, depleted mantle⁴⁴. The depleted-mantle degassing rate today is $^3\text{He} = 1,060 \text{ mol yr}^{-1}$ from refs 1, 42, and increases in the past. Red line, incompletely degassing depleted mantle. The endpoints of the yellow and red lines represent the 'high- $^3\text{He}/^4\text{He}$ ' plume components and present-day MORB, respectively. Details of model parameters and the calculation are in Supplementary Information.

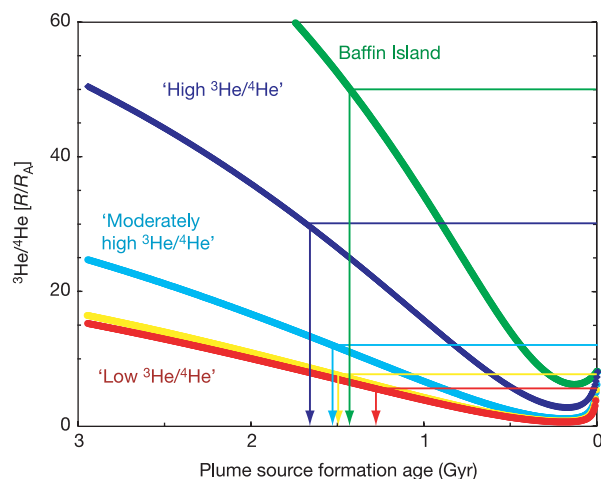


Figure 5 | Plume source formation ages from the incomplete degassing model. Th and U in plume sources are calculated on the assumption that the average Th and Th/U ratio of OIB He groups are formed by 6% partial melting; Baffin endmember with Th = 0.02 p.p.m. and Th/U = 3.31. OIB sources are assumed to represent mixtures of depleted mantle and a constant volume of a recycled component with variable Th + U content (see text). The $^3\text{He}/^4\text{He}$ ratio of a plume source as a function of its formation age is calculated on the basis of the He evolution of the depleted mantle plus the variable contributions of radiogenic ^4He for each He group.

For simplicity, we assume that the amounts of recycled component in all plume sources are constant, and allow the composition (the Th + U content) to vary to calculate the ^4He production rate of the recycled components. For each He-isotope group of OIB, the production rate of ^4He and the $^3\text{He}/^4\text{He}$ ratios can be calculated as a function of the plume source formation age. For the input parameters illustrated in Fig. 5, the four $^3\text{He}/^4\text{He}$ groups and the Baffin Island endmember yield formation ages of between about 1.3 and 1.7 Gyr, which is consistent with previously suggested plume source ages (see ref. 43, for example). Of course, the derived ages vary depending on the choice of model parameters, but the small range of formation ages for all of the $^3\text{He}/^4\text{He}$ groups of OIB is a robust result for any set of inputs. The mantle $^3\text{He}/^4\text{He}$ evolution as presented here is consistent with the whole mantle being 'depleted' by continent and ocean crust formation. However, the addition of recycled crustal components to plume sources, combined with their ancient formation ages, results in incompatible element 'enriched' plume sources, isolated from the depleted mantle, some of which evolve to low $^3\text{He}/^4\text{He}$ and Nd isotope ratios and high Sr and Pb isotope ratios.

Thus, the new global compilation of oceanic basalt data reveals striking similarities between high- $^3\text{He}/^4\text{He}$ OIB and MORB in terms of their trace-element and Pb-Nd-Sr isotope ratios. In addition, global OIB show systematic variability of incompatible element abundances and He-Pb-Sr-Nd isotope ratios, which reflect variable contributions from recycled crust. A model of incomplete degassing for the mantle that allows for the isolation of plume sources reproduces the $^3\text{He}/^4\text{He}$ ratios of MORB and OIB. We conclude that the helium isotope evolution of the mantle does not require the existence of a primordial, undegassed high- $^3\text{He}/^4\text{He}$ mantle reservoir but rather the high- $^3\text{He}/^4\text{He}$ component of OIB represents mantle that has been subject to depletion in incompatible elements by the formation of continents and ocean crust throughout Earth history.

Received 1 December 2004; accepted 14 June 2005.

- Lupton, J. E. & Craig, H. Excess ^3He in oceanic basalts; evidence for terrestrial primordial helium. *Earth Planet. Sci. Lett.* **26**, 133–139 (1975).
- Kurz, M. D., Jenkins, W. J. & Hart, S. R. Helium isotopic systematics of oceanic islands and mantle heterogeneity. *Nature* **297**, 43–47 (1982).
- Hart, S. R., Hauri, E. H., Oschmann, L. A. & Whitehead, J. A. Mantle plumes and

- entrainment: Isotopic evidence. *Science* **256**, 517–520 (1992).
- Farley, K. A., Natland, J. H. & Craig, H. Binary mixing of enriched and undegassed (primitive?) mantle components (He, Sr, Nd, Pb) in Samoan lavas. *Earth Planet. Sci. Lett.* **111**, 183–199 (1992).
- O'Nions, R. K., Evensen, N. M. & Hamilton, P. J. Geochemical modeling of mantle differentiation and crustal growth. *J. Geophys. Res.* **84**, 6091–6101 (1979).
- Allègre, C. J., Staudacher, T. & Sarda, P. Rare gas systematics: formation of the atmosphere, evolution and structure of the earth's mantle. *Earth Planet. Sci. Lett.* **81**, 127–150 (1987).
- Graham, D. W. in *Noble Gases in Geochemistry and Cosmochemistry* (eds Porcelli, D., Ballentine, C. J. & Wieler, R.) 247–317 (Mineralogical Society of America, Washington DC, 2002).
- Kurz, M. D., Jenkins, W. J., Hart, S. R. & Clague, D. Helium isotopic variations in volcanic rocks from Loihi Seamount and the Island of Hawaii. *Earth Planet. Sci. Lett.* **66**, 388–406 (1983).
- Graham, D. W., Humphris, S. E., Jenkins, W. J. & Kurz, M. D. Helium isotope geochemistry of some volcanic rocks from Saint Helena. *Earth Planet. Sci. Lett.* **110**, 121–131 (1992).
- Hanyu, T. & Kaneoka, I. The uniform and low $^3\text{He}/^4\text{He}$ ratios of HIMU basalts as evidence for their origin as recycled materials. *Nature* **390**, 273–276 (1997).
- Moreira, M. & Kurz, M. D. Subducted oceanic lithosphere and the origin of the 'high mu' basalt helium isotopic signature. *Earth Planet. Sci. Lett.* **189**, 49–57 (2001).
- Stuart, F. M., Lass-Evans, S., Fitton, J. G. & Ellam, R. M. High $^3\text{He}/^4\text{He}$ ratios in picritic basalts from Baffin Island and the role of a mixed reservoir in mantle plumes. *Nature* **424**, 57–59 (2003).
- Hofmann, A. W., Jochum, K. P., Seufert, M. & White, W. M. Nb and Pb in oceanic basalts: new constraints on mantle evolution. *Earth Planet. Sci. Lett.* **79**, 33–45 (1986).
- Zindler, A. & Hart, S. R. Helium: problematic primordial signals. *Earth Planet. Sci. Lett.* **79**, 1–8 (1986).
- Hofmann, A. W. Mantle geochemistry: the message from oceanic volcanism. *Nature* **385**, 219–229 (1997).
- Coltice, N. & Ricard, Y. Geochemical observations and one layer mantle convection. *Earth Planet. Sci. Lett.* **174**, 125–137 (1999).
- Creager, K. C. & Jordan, T. H. Slab penetration into the lower mantle. *J. Geophys. Res.* **89**, 3031–3049 (1984).
- Dziewonski, A. M. & Woodhouse, J. H. Global images of the Earth's interior. *Science* **236**, 37–48 (1987).
- Grand, S. P. & van der Hilst, R. D. Global seismic tomography: A snapshot of convection in the Earth. *GSA Today* **7**, 1–7 (1997).
- van der Hilst, R. D. & Karason, H. Compositional heterogeneity in the bottom 1000 kilometers of Earth's mantle: Toward a hybrid convection model. *Science* **283**, 1885–1888 (1999).
- Abedini, A. A. & Hurwitz, S. A global dataset of noble gas concentrations and their isotopic ratios in volcanic areas. *Eos Trans. AGU* **85**, abstract V51B–0524 (2004).
- O'Nions, R. K. & Oxburgh, E. R. Heat and helium in the Earth. *Nature* **306**, 429–431 (1983).
- Mukhopadhyay, S., Lassiter, J. C., Farley, K. A. & Bogue, S. W. Geochemistry of Kauai shield-stage lavas: Implications for the chemical evolution of the Hawaiian plume. *Geochim. Geophys. Geosyst.* **4**, doi:10.1029/2002GC000342 (2003).
- Farley, K. A. Rapid cycling of subducted sediments into the Samoan mantle plume. *Geology* **23**, 531–534 (1995).
- Hofmann, A. W. & White, W. M. Mantle plumes from ancient oceanic crust. *Earth Planet. Sci. Lett.* **57**, 421–436 (1982).
- Craig, H., Clarke, W. B. & Beg, M. A. Excess ^3He in deep water on the East Pacific Rise. *Earth Planet. Sci. Lett.* **26**, 125–132 (1975).
- Hauri, E. H., Lassiter, J. C. & DePaolo, D. J. Osmium isotope systematics of drilled lavas from Mauna Loa, Hawaii. *J. Geophys. Res.* **101**, 11793–11806 (1996).
- Kellogg, L. H. & Wasserburg, G. J. The role of plumes in mantle helium flux. *Earth Planet. Sci. Lett.* **99**, 276–289 (1990).
- Porcelli, D. & Wasserburg, G. J. Mass transfer of helium, neon, argon, and xenon through a steady-state upper mantle. *Geochim. Cosmochim. Acta* **59**, 4921–4937 (1995).
- Anderson, D. L. A model to explain the various paradoxes associated with mantle noble gas geochemistry. *Proc. Natl Acad. Sci. USA* **95**, 9087–9092 (1998).
- Albarède, F. Time-dependent models of U–Th–He and K–Ar evolution and the layering of mantle convection. *Chem. Geol.* **145**, 413–429 (1998).
- Montelli, R. et al. Finite-frequency tomography reveals a variety of plumes in the mantle. *Science* **303**, 338–343 (2004).
- Morgan, J. P. & Morgan, J. W. Two-stage melting and the geochemical evolution of the mantle: a recipe for mantle plum-pudding. *Earth Planet. Sci. Lett.* **170**, 215–239 (1999).
- Ballentine, C. J., van Keken, P. E., Porcelli, D. & Hauri, E. H. Numerical models, geochemistry and the zero-paradox noble-gas mantle. *Phil. Trans. R. Soc. Lond. A* **360**, 2611–2631 (2002).
- Allègre, C. J., Hofmann, A. W. & O'Nions, R. K. The argon constraints on mantle structure. *Geophys. Res. Lett.* **23**, 3555–3557 (1996).

36. Davies, G. F. Geophysically constrained mantle mass flows and the ^{40}Ar budget: A degassed lower mantle? *Earth Planet. Sci. Lett.* **166**, 149–162 (1999).
37. Lassiter, J. C. Role of recycled oceanic crust in the potassium and argon budget of the Earth: Toward a resolution of the 'missing argon' problem. *Geochem. Geophys. Geosyst.* **5**, doi:10.1029/2004GC000711 (2004).
38. Wood, B. J. & Blundy, J. D. The effect of cation charge on crystal-melt partitioning of trace elements. *Earth Planet. Sci. Lett.* **188**, 59–71 (2001).
39. Parman, S. W., Kurz, M. D., Hart, S. R. & Grove, T. L. Solubility of helium in olivine at 1 atmosphere. *Eos Trans. AGU* **85**, abstract U41A–0725 (2004).
40. Staudacher, T. & Allègre, C. J. Terrestrial xenology. *Earth Planet. Sci. Lett.* **60**, 389–406 (1982).
41. Goldstein, S. L., O'Nions, R. K. & Hamilton, P. J. A Sm–Nd isotopic study of atmospheric dusts and particulates from major river systems. *Earth Planet. Sci. Lett.* **70**, 221–236 (1984).
42. Farley, K. A., Maierreimer, E., Schlosser, P. & Broecker, W. S. Constraints on mantle He-3 fluxes and deep-sea circulation from an oceanic general circulation model. *J. Geophys. Res. Solid Earth* **100**, 3829–3839 (1995).
43. Hart, S. R. A large scale isotope anomaly in the southern hemisphere mantle. *Nature* **309**, 753–757 (1984).
44. Salters, V. J. M. & Stracke, A. Composition of the depleted mantle. *Geochem. Geophys. Geosyst.* **5**, doi:10.1029/2003GC000597 (2004).
45. Sun, S.-S. & McDonough, W. F. in *Magmatism in the Ocean Basins* (eds Saunders, A. D. & Norry, M. J.) 313–345 (Blackwell Scientific, Oxford/Boston, 1989).
46. Harper, C. L. & Jacobsen, S. B. Noble gases and Earth's accretion. *Science* **273**, 1814–1818 (1996).
47. Porcelli, D. & Ballentine, C. J. in *Noble Gases in Geochemistry and Cosmochemistry* (eds Porcelli, D., Ballentine, C. J. & Wieler, R.) 411–480 (Mineralogical Society of America, Washington DC, 2002).
48. Tolstikhin, I. & Hofmann, A. W. Early crust on top of the Earth's core. *Phys. Earth Planet. Inter.* **148**, 109–130 (2004).
49. Lehnert, K., Su, Y., Langmuir, C., Sarbas, B. & Nohl, U. A global geochemical database structure for rocks. *Geochem. Geophys. Geosyst.* **1**, doi:10.1029/1999GC000026 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Abedini for sharing her noble gas database; W. White and D. Graham for comments that helped to improve the manuscript; A.W. Hofmann and A. Class for discussions. This study was supported by National Science Foundation grants. This is LDEO Contribution number 6791.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.C. (class@ldeo.columbia.edu).

Myosin domain evolution and the primary divergence of eukaryotes

Thomas A. Richards^{1,2†} & Thomas Cavalier-Smith²

Eukaryotic cells have two contrasting cytoskeletal and ciliary organizations. The simplest involves a single cilium-bearing centriole, nucleating a cone of individual microtubules (probably ancestral for unikonts: animals, fungi, Choanozoa and Amoebozoa). In contrast, bikonts (plants, chromists and all other protozoa) were ancestrally biciliate with a younger anterior cilium, converted every cell cycle into a dissimilar posterior cilium and multiple ciliary roots of microtubule bands. Here we show by comparative genomic analysis that this fundamental cellular dichotomy also involves different myosin molecular motors. We found 37 different protein domain combinations, often lineage-specific, and many previously unidentified. The sequence phylogeny and taxonomic distribution of myosin domain combinations identified five innovations that strongly support unikont monophyly and the primary bikont/unikont bifurcation. We conclude that the eukaryotic cenancestor (last common ancestor) had a cilium, mitochondria, pseudopodia, and myosins with three contrasting domain combinations and putative functions.

Myosins bind to actin, hydrolysing ATP to produce physical force, and are fundamental in eukaryotic cytokinesis, organellar transport, cell polarization, intracellular transport and signal transduction^{1,2}. They evolved, like microtubules, during the origin of eukaryotes³. Their head domains, containing the ATPase and actin-binding activities, are connected to a range of amino-terminal and carboxy-terminal domains⁴, corresponding to the variety of molecular cargos that myosins bind and move. Sequence phylogeny and protein domain combinations have previously been used to establish 18 myosin 'classes'^{4–7}, although additional myosin types have also been reported⁸. The function of many myosin 'classes' has been characterized and is distinct, but full functional properties are unknown⁴ for others or for currently unclassified myosins. Myosin and the related kinesin⁹ gene families along with protein-synthesis elongation factors form the TRAFAC class of the P-loop GTPases that originated by the deletion of strands 6 and 7 in the GTPase core and the addition of two N-terminal strands¹⁰.

Studying the diversification of eukaryote-specific molecular motors that interact closely with the cytoskeleton may be particularly fruitful for understanding phylogenetic patterns, the cellular apparatus and the functional attributes of early eukaryotes. Gene families with numerous paralogues (genes related by duplication but with non-identical functions), such as myosins, are often considered unhelpful for reconstructing ancient evolutionary relationships because of their very complexity. However, with sufficient taxon sampling and reliable sequence phylogenies, patterns of sequence synapomorphies (derived character states shared by two or more taxa) and paralogue distribution can be used to map ancient evolution. No myosin has been found in prokaryotes; thus, an innovative shift in nucleotide-binding specificity (GTP to ATP) occurred to form the myosin–kinesin ancestor at the very origin of eukaryotes—in which actin and tubulin cytoskeletons had a central role³. Because both myosins and kinesins underwent marked diversification and domain rearrangements, the comparative study of these molecular motors offers great potential for disentangling early eukaryote evolution. Here we show that there are more than twice as

many myosin types as previously described, all possessing unique domain structures and/or arrangements. This diversity can be divided into a limited number of subfamilies, of which three were present in the eukaryote cenancestor. Several features of myosin diversification strongly support a primary eukaryotic unikont/bikont bifurcation^{3,11–13}.

Immense diversity of myosin types

Our survey of the myosin gene family revealed 37 myosin types with different combinations of protein domains and scattered taxonomic distribution (Fig. 1a). The diversity of myosin paralogues encoded by each eukaryote varies considerably; for example, *Phytophthora ramorum* has 25 myosin genes encoding 13 different types, and humans have 12 (Fig. 1a), 6 of which are also present in *Dictyostelium*. The myosin types in *Phytophthora* and humans (Fig. 1a) represent independent peaks in evolutionary diversity of this gene family. In contrast, no myosin head domains could be identified from the flagellates *Giardia intestinalis*, *Trichomonas vaginalis* or the red alga *Cyanidioschyzon merolae*¹⁴ with either BLASTp or PSI-BLAST¹⁵. The diversity of myosin types reported here indicates that many more might await discovery. Thirty myosin types were specific to narrow evolutionary lineages; for example, type 32 with multiple N-terminal WD40 domains was found only in Apicomplexa.

Multiple domain losses and gains

Myosin phylogeny reveals many instances of domain loss by deletion or divergence; for example, type 27 myosins (Fig. 1a) with no identified tail domains are clearly nested on the tree within clades comprising molecules with distinctive tail domains (Fig. 2), indicating a polyphyletic origin. Even paralogues not thus positioned have restricted taxonomic distribution, indicating that they might have arisen by recent tail loss. It is therefore possible that all ancestral myosins had long tails and that myosins with no identifiable tail domain (type 27) arose secondarily by multiple independent simplifications, making type 27 an artificial category.

An example of domain loss is type 27 of *Chlamydomonas*, which

¹Department of Zoology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK. ²Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK. †Present address: School of Biological and Chemical Sciences, University of Exeter, Washington Singer Laboratories, Perry Road, Exeter EX4 4QG, UK.

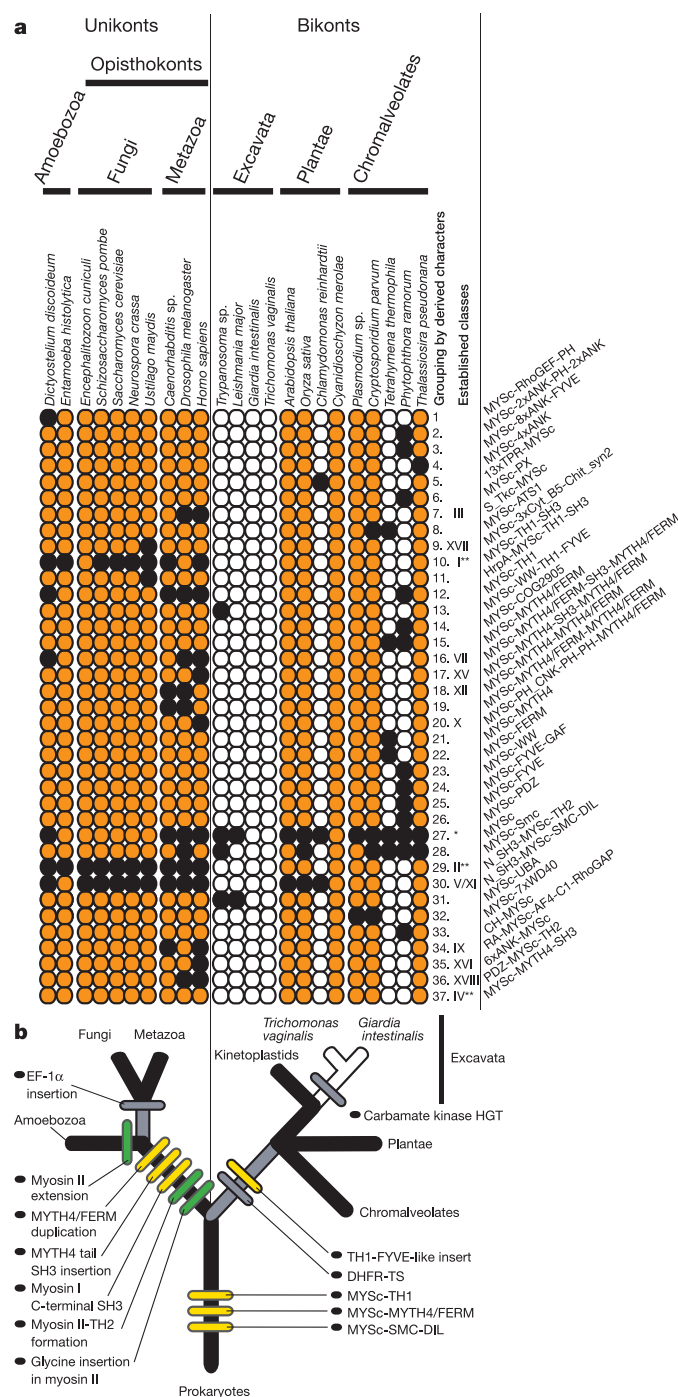


Figure 1 | Taxonomic distribution and evolutionary history of myosin paralogues. **a**, Comparative genomic survey of myosin paralogues in 23 key eukaryotic taxa. Taxa are shown on the X-axis and myosin domain order and classification on the Y-axis. Black dots indicate detection; open dots indicate no data available; orange dots indicate absence from a completed genome project. All myosins identified are listed with accession numbers in Supplementary Table 1. The type numbers should not be confused with previous class designations (for example ref. 4); equivalents are given in the right-hand column and in the text where appropriate. To conserve space, classes VI/VIII/XIV/XIII are indicated with an asterisk. Double asterisks, also detected in *Acanthamoeba castellanii*. **b**, Schematic tree showing myosin-derived synapomorphies (green bars) and three previously published shared characters (grey bars)^{11,40,41}. Yellow bars indicate synapomorphy plus secondary loss. MYSc, myosin head domain.

strongly groups within type 30 (class XI) myosins and therefore is likely to have lost the C-terminal part of the molecule containing the dilute (DIL) domain found in all other members of the clade to which it belongs (Fig. 2). Although we detected no coiled-coil (Smc) domain in this molecule, it has a peptide tail that in its closest relatives carries an Smc domain, indicating that this region might have diverged beyond recognition. It is also clear that much variety in myosin domain organization arose by secondary domain losses, either by partial deletion or by divergence: current bioinformatic methods cannot distinguish absence from extreme divergence. The tree reveals several clear examples of novel domain gains, for example chitin synthetase to generate fungal class XVII proteins (type 9) and COG2905 yielding a *Phytophthora* myosin (type 14).

Three ancient myosin subfamilies

The 23 completed or near-complete genomes surveyed belong to five higher taxonomic units, namely opisthokonts, Excavata, Plantae, chromalveolates and Amoebozoa (Fig. 1), covering five of the six known eukaryotic supergroups^{13,16,17}, with only Rhizaria currently unsampled. Only 7 of the 37 myosin arrangements are found in more than one supergroup; most evolved after early eukaryote diversification. If we allow for fusions, partial deletions, duplications, and losses, we can use shared derived characters to rationalize myosin diversity into five broad ancestral myosin subfamilies. On the basis of taxonomic distribution, three of these seem to have been present in the eukaryote ancestor (Fig. 1b). Of the other two, unikont-specific myosin II (type 29) is phylogenetically well defined, whereas a large weakly resolved group of chromalveolate myosins with a range of different C-terminal domains constitutes the second non-ancestral 'subfamily' (Fig. 2).

The broad taxonomic distribution of myosins with coiled-coil and dilute domains (MYSc-SMC-DIL, classes V and XI; here called MSD subfamily; type 30; Fig. 1a) in Plantae, opisthokonts and Amoebozoa, and their grouping in two robust clades (which we cannot exclude from being a single clade; Fig. 2) indicates that this arrangement might have arisen in the ancestral eukaryote and was lost by excavates and chromalveolates. The presence of an N-terminal SH3 domain varies between members of the MSD subfamily (Fig. 2). These domains are structurally similar to other SH3 domains but have many sequence differences. SH3 domains have conserved structures¹⁸ but very variable sequences, which can make them difficult to identify; sequence alignments indicate that many MSD and class II myosins might possess an N-terminal SH3 domain not identified by conserved domain database (CDD) searches (Supplementary Fig. 1). The simplest interpretation of the scattered phylogenetic distribution of this domain within these myosin types is a combination of secondary losses and sequence divergence from an ancestral myosin that possessed an N-terminal SH3 domain. Alternative explanations involving independent additions, although possible given the numerous incidences of recombination involving SH3 domains, are less parsimonious. Thus, the ancestral eukaryote probably had a myosin with domain structure N-SH3-MYSc-SMC-DIL (Fig. 1b).

The second putatively ancestral subfamily comprises myosins of classes IV/VII/XII/XV (types 16–18 and 37) with an MYTH4/FERM domain. They are found in animals, Amoebozoa and chromalveolates, indicating that a myosin gene with one MYTH4/FERM domain might have been present in the eukaryotic ancestor before undergoing multiple secondary losses or gene modifications. Alternatively, these domains might instead have become associated on separate occasions; the presence of a MYTH4/FERM domain at the N terminus of plant kinesins (for example GenBank accession number CAE03597) indicates that these domains might have recombined among distantly related molecular motors at least once.

Animals and Amoebozoa alone have MYTH4/FERM plus SH3 domain tails (Supplementary Fig. 2), representing a synapomorphy for unikont holophyly (being a monophyletic group with a single

evolutionary origin and including all descendents of its cenancestor). A second potential unikont synapomorphy is the duplication of the MYTH4/FERM region, indicating that the ancestral unikont domain structure might have been MYSc-MYTH4/FERM-SH3-MYTH4/FERM (type 16—myosin VII), as this is found in metazoa and *Dictyostelium* (although not in fungi, probably because of gene

loss); simpler tail structures in this subfamily can readily be derived from this by differential partial deletions and/or domain insertions (for example, PH domains in vertebrate class X myosins; type 20). MYTH4/FERM myosins are dispersed among three clades (Fig. 2), but the tree base is too weakly resolved to disprove holophyly. A single origin of MYTH4/FERM tails in the cenancestor is the most

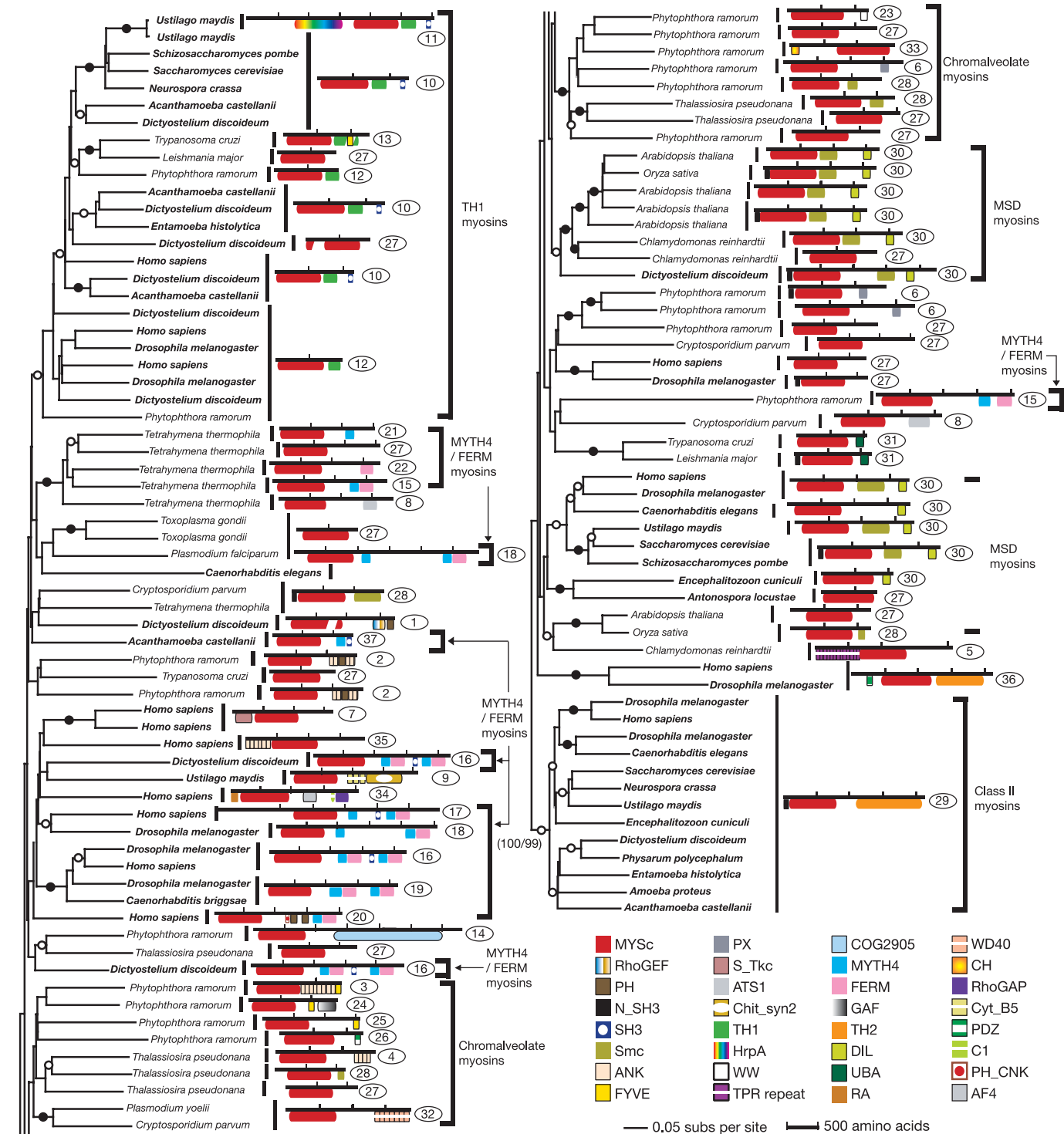


Figure 2 | Myosin head domain phylogeny (bayesian consensus: 118 myosins; 357 characters). Circled 1–37 designate domain combinations (Fig. 1a). Domains are labelled to scale (500 amino-acid residues indicated); myosin head domains (MYSc) shown in red, see key for colour-coding and Supplementary Table 1 for other abbreviations and naming. Square brackets label the five myosin subfamilies proposed here. Support values (bayesian

posterior probability, 1,000 maximum-likelihood distance bootstraps or 100 Protpars bootstraps) are marked if all are more than 90% (filled circle) or more than 60% (open circle). Support values in brackets are from separate sequence-rich distance and parsimony analyses with long branches excluded. Unikonts are shown in bold. Bootstrap values over 50%, accession numbers and phylogenetic methods are given in Supplementary Fig. 4.

parsimonious interpretation, followed by differential duplication and domain losses and gains. Moreover, two properties of this subfamily—the MYTH4/FERM duplication and the SH3 insertion—support unikont holophyly (Fig. 1b).

The third probably ancestral subfamily is myosin I (types 10–12) with a membrane-binding TH1 domain tail. It is found in excavates, chromalveolates, opisthokonts and Amoebozoa but not in Plantae, implying that it was cenacestral but lost by Plantae. It is unlikely that Plantae diverged before myosin I formation, because the *dhfr-ts* fusion and their pattern of ciliary transformation support the inclusion of Plantae within bikonts^{11–13}. The addition of an SH3 domain to the C terminus of this protein was detected only in unikonts, for which it may be synapomorphic (type 10; Fig. 1b and Supplementary Fig. 3). Phylogenetic analyses did not resolve this portion of the tree with significant support but did not clearly contradict this inference. In addition, bikont myosin type 12 genes contained a highly variable approximately 60 amino-acid-residue insertion within the TH1 domain. This is identified as a FYVE protein domain in trypanosomes but is unidentifiable in *Phytophthora*. Amino-acid alignments reveal that, although highly variable, this insertion contains several conserved positions, including four cysteine residues and VRV and KST motifs, indicating that it might be homologous (Supplementary Fig. 3) and a synapomorphy for bikonts or a subset of them (Fig. 1b).

Interestingly, both myosin I and MYTH4/FERM myosins have SH3 domains in their tail in unikonts but not in bikonts (Fig. 1a, b), whereas the third putatively cenacestral myosin subfamily (MSD) has an N-terminal-type SH3 domain. Because the three subfamilies must have arisen by two successive gene duplications of the first myosin gene in the cenacestor, the cenacestral myosin might have had an SH3-related domain; tandem duplication, differential deletions and divergence could have quickly generated each of the three primary myosin subfamilies. By this model the absence of SH3 domains from bikont class 1 (type 12) and MYTH4/FERM myosins would be secondary losses and thus synapomorphies for

bikonts rather than for unikonts. However, the SH3 domain is in so many eukaryotic proteins that it must have been highly mobile (or have originated independently; that is, polyphyletically) during early eukaryote evolution. Consequently the insertion of SH3 domains into tails of an early unikont myosin or myosins is plausible.

Myosins and the unikont/bikont split

Class II myosins (Fig. 1) were proposed together with class I myosins to be the most ancient of all⁴; above we argued that MSD, MYTH4/FERM, and class I myosins, all occur in the widest diversity of eukaryotic supergroups and are therefore likely to be ancestral. The absence of myosin II from bikonts (Fig. 1), and the significant bootstrap support for its holophyly (Fig. 2), indicate that it might not have been a cenacestral myosin but instead a synapomorphy for unikonts only (Fig. 1a, b). A novel glycine residue inserted at position 507 (*Dictyostelium discoideum*) (Fig. 3) within all class II-derived myosins only (except myosin class XVIII—type 36) unambiguously supports the holophyly and derived nature of this paralogue. In some genes, indels can be ambiguous characters because of alignment uncertainty or evidence of multiple changes at the same site in different taxa¹⁹; such complications are absent in this case and the insertion is the derived state. This character, the strongly supported monophyly of myosin II (Fig. 2) and the unique myosin II coiled-coil tail (TH2) domain all make it highly improbable that the insertion occurred more than once. The less parsimonious possibility exists that myosin II was present in the first eukaryotes but was lost by the common ancestor of all sampled bikonts (by deletion or extreme divergence). Although some myosins show evidence of secondary loss and/or extreme divergence, there is no evidence of either for myosin II in the ten unikont species sampled. Characterizing myosins from Rhizaria would test our interpretation, which would be simply disproved if any have myosin II. Although the sequence phylogeny is unresolved for myosins with the MYTH4/FERM duplication and the acquisition of SH3 by class I and MYTH4/FERM myosins, all three are synapomorphies supporting unikont holophyly; thus, five independent synapomorphies give the same answer (Fig. 1b). Postulating four independent secondary losses of these paralogues is unparsimonious.

The apparent absence of myosin head domains in the two metamonad genomes (*Giardia* and *Trichomonas*) and the red alga *Cyanidioschyzon merolae* indicates that these organisms might lack myosins or that the myosins have evolved so radically that they are currently unidentifiable. Extreme sequence evolution in the common ancestor of all bikonts, and/or gene loss, could have masked the existence of bikont orthologues of the unikont myosin synapomorphies. Such hypothetical alternatives would be synapomorphies for bikonts. Independent later losses in all bikont lineages sampled would be even less parsimonious. Thus, either way, the myosin synapomorphy distribution data (Fig. 1) collectively support the partition of eukaryotes into unikonts and bikonts and are consistent with the holophyly of both groups¹². *Giardia* and *Trichomonas* protein-encoding genes are notoriously fast-evolving compared with those of most other eukaryotes (except microsporidia²⁰), which might explain why we found no myosins in their genomes. Significantly, we detected myosin class II (type 29) and class XI (type 30) in microsporidia, whose genes generally evolve even faster than those of metamonads; this ready detectability in the remarkably fast-evolving microsporidian genome makes secondary 'losses' in bikonts through rapid divergence unlikely, especially as myosin II is uniformly absent from plants and chromalveolates, which do not show unusually rapid divergence in their protein-encoding genes. It is therefore more likely that the five unikont myosin synapomorphies arose after a primary divergence of eukaryotes into unikonts and bikonts than that all five were ancestrally lost by bikonts. Together they provide the best available evidence for the holophyly of unikonts.

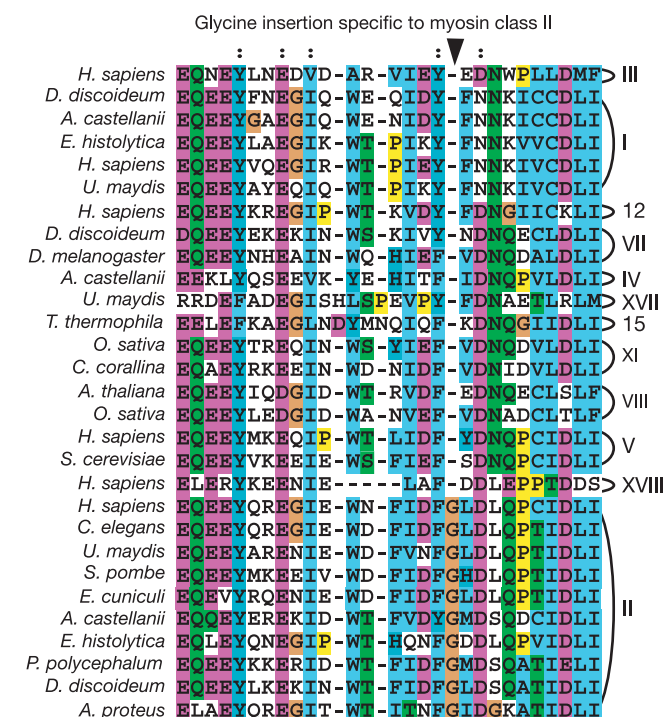


Figure 3 | Section from a myosin sequence alignment, including a representative selection of myosin types. The section illustrates a glycine insertion specific to myosin class II genes. Myosin classifications/types (Fig. 1a) are shown at the right.

Our trees (if appropriately rooted: the root shown is arbitrary) weakly support the idea that class II and MSD myosins are related⁴. TH2 domains of class II myosins contain multiple sequence regions forming heptad repeats with similarity to Smc and other defined protein domains known to have a function in forming coil-coiled structures. We constructed an amino-acid alignment to investigate the potential homology between TH2 myosins and paralogues that possess Smc-type tails and/or a DIL-type tail domain (Supplementary Fig. 1). This shows that the tails are highly variable but have conserved character blocks, many present in the DIL-type tail (MSD myosins; class V and XI—type 30; Fig. 1a) or the TH2 domain (myosin II—type 29), but not both, establishing that myosin II TH2 domains are unique. Some regions do show weak homology between class II and MSD genes, consistent with a common ancestry and radical sequence divergence. The existence of TH2 domains unconnected to myosin head domains (for example *Giardia intestinalis* (GenBank accession number EAA38371) *Oryza sativa* (NP_921528) and several other eukaryotes) or connected to a kinesin domain in fungi (GenBank accession number T51930) indicates that it might have undergone illegitimate domain recombination at least once. We therefore cannot exclude the possibility that myosin II arose by gene fusion rather than from simple gross divergence of the tail from an MSD ancestor, but its support for unikont holophyly does not depend on its precise mode of origin.

Nature of the ancestral eukaryote

The presence of myosin II and its conserved amino-acid insertion in five diverse Amoebozoa (*Acanthamoeba castellanii*, *Physarum polycephalum* and *Dictyostelium discoideum*, *Amoeba proteus*—a member of Lobosea (naked aerobic amoebae with broad finger-like pseudopods but no cilia or ciliary root apparatus)—and *Entamoeba histolytica*, representing the amitochondrial Archamoebae formerly postulated to be early branching eukaryotes²¹) supports the arguments¹² that the unikont–bikont bifurcation is the oldest evolutionary diversification of known eukaryotes and shows that all these Amoebozoa are unikonts. The myosin II tree has 87%/62% bootstrap support for amoebozoan holophyly; previously Amoebozoa might have been paraphyletic, occupying a basal position to all eukaryotes, with some representatives (such as *Dictyostelium*) closer to opisthokonts, others closer to bikonts, and others diverging before the unikont/bikont bifurcation²². Monophyly of Amoebozoa has only weak to moderate bootstrap support in most 18S rRNA phylogenies^{22–25}. Small subsets of amoebozoan taxa consistently form monophyletic clusters in sequence phylogenies of numerous nuclear or mitochondrial proteins^{26–28}, but taxon sampling is far too narrow to demonstrate holophyly or rooting of the Amoebozoa. Because Lobosea are entirely devoid of cilia, unlike Myxogastrea and Variosea²², it might be argued (given the poor resolution of all sequence trees that include Lobosea) that the absence of cilia could be a primitive character and Lobosea might be the deepest branch in the eukaryotic tree, branching before cilia evolved²². The distribution of the myosin II synapomorphies makes this unlikely and strongly implies that ancestors of Lobosea lost cilia^{3,13,22–24}. Amoebozoa are monophyletic (87%/62% bootstrap support) in the myosin II phylogeny, making paraphyly of Amoebozoa with respect to opisthokonts unlikely; the presence of a homologous approximately 130-residue extension to myosin II in all Amoebozoa analysed, but in no other eukaryotes (Supplementary Fig. 1), also supports amoebozoan holophyly. Although one key amoebozoan group (Discosea²²) awaits sampling, there is no reason to suspect that it diverged before the fundamental unikont–bikont split or before the split between opisthokonts and the Amoebozoa sampled here. Monophyly of Mycetozoa (for example *Dictyostelium* and *Physarum*), previously unclear^{22,23}, was recovered with 85%/63% bootstrap support. Although the concept of a basal eukaryote bifurcation between unikonts and bikonts is relatively new¹², a recent comprehensive bayesian analysis of 18S rRNA shows a clear

bipartition between unikonts and bikonts, and amoebozoan holophyly, both well supported²⁹—unlike earlier distance trees¹⁶.

Patterns of pseudopodial shape and movement seem very different between Amoebozoa and Rhizaria^{13,30}, which both include numerous amoeboid lineages. Amoebozoa have flat or lobose pseudopods²², whereas Rhizaria tend to have thread-like filopodia or anastomosing reticulopodia—which is consistent with these two groups' being unrelated^{13–31}. Their membership of unikonts and bikonts, respectively, indicates that the formation of pseudopodia in eukaryotic cells was probably an attribute of the last common eukaryotic ancestor. If myosin II, which functions in cytokinesis in unikonts, was genuinely absent from the ancestral eukaryote this function must originally have been performed by a different myosin, possibly the related MSD myosins. The numerous types of myosin in chromalveolates with novel domain organizations might be related to analogous functional replacements of the absent MSD myosins.

Discussion

The five new myosin synapomorphies for unikonts pinpoint the root of the eukaryotic tree with greater confidence and precision; they mean that the finding of a triple gene fusion, originally used to support unikont holophyly¹², in a red alga¹⁴ (a bikont) does not invalidate the concept of unikont holophyly. Establishing the root position allows us to specify several key features of the last common ancestral eukaryote cell: an endosymbiont-derived mitochondrion, a cilium and centriole (most parsimoniously a single one with a cone of root microtubules^{21,22}), and the cellular machinery to form pseudopodia. The amoebozoan flagellate *Phalansterium* with all these characters may be the best extant model for the ancestral eukaryotic phenotype^{3,22}. As argued above, the cenacestral eukaryote probably had three different myosins with contrasting tail domains: myosin I, MYTH4/FERM myosins and MSD myosins. How the primary functions of myosin in cytokinesis, phagocytosis, pseudopodial and vesicle movement—all central to the life of the first eukaryote cells—were partitioned between these myosins cannot be clearly inferred from present data. The recent demonstration of an essential function for MYTH4/FERM myosin in *Dictyostelium* adhesion, important in both phagocytosis³² and motility³³, indicates that this might have been its early function; its loss in both Plantae and Fungi, which independently evolved cell walls—thus losing both phagocytosis and amoeboid motility—is consistent with this. It is therefore tempting to indicate that MSD myosins might ancestrally have been responsible for cytokinesis (a role retained by their myosin II descendants) and pseudopodial activity. Functional studies of a more phylogenetically representative set of the myosins detailed here are needed to test this and to clarify major evolutionary shifts in myosin function. Physiological and genomic studies of myosin function and diversity in bikonts is especially needed (particularly Rhizaria, excavates, chromalveolates and lower plants). Given the marked differences in pseudopodial organization in Amoebozoa and Rhizaria, it would be particularly valuable to determine which myosin paralogues are present in Rhizaria; this might reveal novel lineage-specific paralogues, test our tentative conclusion that only three myosin subfamilies were ancestral for all eukaryotes, and yield further improvements in myosin classification.

METHODS

Comparative genome analyses. BLASTp searches obtained all recognizable myosin paralogues from 23 eukaryotic genome projects (up to April 2005; listed in Fig. 1a) by using GenBank eukaryote genome and non-redundant (nr) databases, dictybase, The Institute for Genomic Research (TIGR), Department of Energy Joint Genome Resource and the *Cyanidioschyzon merolae* genome project. Each myosin was then searched against the protein conserved domain database (CDD)³⁴ and the Pfam HMM³⁵ database to identify and classify protein domains. Protein domain identification is limited by the sensitivity of the search system and the diversity of protein domains in the database. Pfam and CDD were used in combination to increase both the sensitivity and the protein diversity. Every individual myosin type (defined here as a unique combination of protein

domains: 1–37 in Fig. 1a; N-terminal SH3-like and IQ domain characteristics were judged too variable for such typing) was then used for BLASTp searches against the GenBank nr database to identify further homologues from organisms additional to the 23 main taxa. Extra species not in Fig. 1 were surveyed to check for consistency or contradiction with the phylogenetic inferences based on the 23 comprehensively surveyed genomes (Supplementary Table 1). PSI BLAST¹⁵ was used to seek highly divergent myosin head domains in genomes of *Giardia*, *Leishmania*, *Trichomonas* and *Cyanidioschyzon* and used myosin class I and II genes as starting seeds with *Dictyostelium* and *Phytophthora* genome sampling, to inform the PSI BLAST alignment process. PSI BLAST was run for 20 iterations, but gene discovery stopped several iterations before all searches finished.

Sequence alignment and phylogenetic analyses. Amino-acid sequences of the myosin head domains were aligned by ClustalX³⁶ and refined manually with Se-Al. Insertions and sequence characters not alignable with confidence were removed. Alignments sampling extensive diversity were initially analysed and then pared down by removing closely related sequences, while maintaining representative taxonomic and paralogue diversity. For the final phylogenetic trees two alignments were analysed: one sampled 357 conserved amino-acid positions only (to reduce long-branch problems) and 118 taxa, representing known myosin diversity. The second increased sampling (150 sequences, 371 characters) but with some long-branch myosin classes removed (for example myosin classes XVIII and XII). The resulting topologies are generally congruent apart from positions weakly supported in all analyses. Edited alignments were analysed by three methods: first, MrBayes 3 (ref. 37); second, maximum-likelihood distance bootstrap values (from 1,000 replicates) (refs 38, 39, and <http://hades.biochem.dal.ca/Rogerlab/Software/software.html#puzzleboot>); and third, 100 Protpars³⁹ bootstrap replicates; Supplementary Fig. 4 gives details.

Received 9 March; accepted 24 June 2005.

- Sellers, J. R. *Myosins* (Oxford Univ. Press, Oxford, 1999).
- Bahler, M. Are class III and class IX myosins motorized signalling molecules? *Biochim. Biophys. Acta* **1496**, 52–59 (2000).
- Cavalier-Smith, T. The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. *Int. J. Syst. Evol. Microbiol.* **52**, 297–354 (2002).
- Thompson, R. F. & Langford, G. M. Myosin superfamily evolutionary history. *Anat. Rec.* **268**, 276–289 (2002).
- Furusawa, T., Ikawa, S., Yanai, N. & Obinata, M. Isolation of a novel PDZ-containing myosin from hematopoietic supportive bone marrow stromal cell lines. *Biochem. Biophys. Res. Commun.* **270**, 67–75 (2000).
- Goodson, H. V. & Spudich, J. A. Molecular evolution of the myosin family: relationships derived from comparisons of amino acid sequences. *Proc. Natl Acad. Sci. USA* **90**, 659–663 (1993).
- Hodge, T. & Cope, M. J. A myosin family tree. *J. Cell Sci.* **113**, 3353–3354 (2000).
- Berg, J. S., Powell, B. C. & Cheney, R. E. A millennial myosin census. *Mol. Biol. Cell* **12**, 780–794 (2001).
- Kull, F. J., Vale, R. D. & Fletterick, R. J. The case for a common ancestor: kinesin and myosin motor proteins and G proteins. *J. Muscle Res. Cell Motil.* **19**, 877–886 (1998).
- Leipe, D. D., Wolf, Y. I., Koonin, E. V. & Aravind, L. Classification and evolution of P-loop GTPases and related ATPases. *J. Mol. Biol.* **317**, 41–72 (2002).
- Stechmann, A. & Cavalier-Smith, T. Rooting the eukaryote tree by using a derived gene fusion. *Science* **297**, 89–91 (2002).
- Stechmann, A. & Cavalier-Smith, T. The root of the eukaryote tree pinpointed. *Curr. Biol.* **13**, R665–R666 (2003).
- Cavalier-Smith, T. Protist phylogeny and the high-level classification of Protozoa. *Eur. J. Protistol.* **39**, 338–348 (2003).
- Matsuzaki, M. *et al.* Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature* **428**, 653–657 (2004).
- Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
- Cavalier-Smith, T. Only six kingdoms of life. *Proc. R. Soc. Lond. B* **271**, 1251–1262 (2004).
- Simpson, A. G. & Roger, A. J. The real ‘kingdoms’ of eukaryotes. *Curr. Biol.* **14**, R693–R696 (2004).
- D’Aquino, J. A. & Ringe, D. Determinants of the SRC homology domain 3-like fold. *J. Bacteriol.* **185**, 4081–4086 (2003).
- Bapteste, E. & Philippe, H. The potential value of indels as phylogenetic markers: position of trichomonads as a case study. *Mol. Biol. Evol.* **19**, 972–977 (2002).
- Hirt, R. P. *et al.* Microsporidia are related to Fungi: evidence from the largest subunit of RNA polymerase II and other proteins. *Proc. Natl Acad. Sci. USA* **96**, 580–585 (1999).
- Cavalier-Smith, T. Archamoebae: the ancestral eukaryotes? *BioSystems* **25**, 25–38 (1991).
- Cavalier-Smith, T., Chao, E. E. & Oates, B. Molecular phylogeny of Amoebozoa and the evolutionary significance of the unikont *Phalansterium*. *Eur. J. Protistol.* **40**, 21–48 (2004).
- Bolivar, I., Fahrni, J. F., Smirnov, A. & Pawlowski, J. SSU rRNA-based phylogenetic position of the genera *Amoeba* and *Chaos* (Lobosea, Gymnamoebia): the origin of gymnamoebae revisited. *Mol. Biol. Evol.* **18**, 2306–2314 (2001).
- Kudryavtsev, A. A., Bernhardt, D., Schlegel, M., Chao, E. E. & Cavalier-Smith, T. 18S ribosomal RNA gene sequences of *Cochliopodium* (Himantismenida) and the phylogeny of Amoebozoa. *Protist* **156**, 215–224 (2005).
- Milyutina, I. A., Aleshin, V. V., Mikrjukov, K. A., Kedrova, O. S. & Petrov, N. B. The unusually long small subunit ribosomal RNA gene found in amitochondriate amoeboid flagellate *Pelomyxa palustris*: its rRNA predicted secondary structure and phylogenetic implication. *Gene* **272**, 131–139 (2001).
- Dacks, J. B., Marinets, A., Doolittle, W. F., Cavalier-Smith, T. & Logsdon, J. M. Jr Analyses of RNA polymerase II genes from free-living protists: phylogeny, long branch attraction, and the eukaryotic big bang. *Mol. Biol. Evol.* **19**, 830–840 (2002).
- Bapteste, E. *et al.* The analysis of 100 genes supports the grouping of three highly divergent amoebae: *Dictyostelium*, *Entamoeba*, and *Mastigamoeba*. *Proc. Natl Acad. Sci. USA* **99**, 1414–1419 (2002).
- Lang, B. F., O’Kelly, C., Nerad, T., Gray, M. W. & Burger, G. The closest unicellular relatives of animals. *Curr. Biol.* **12**, 1773–1778 (2002).
- Berney, C., Fahrni, J. & Pawlowski, J. How many novel eukaryotic ‘kingdoms’? Pitfalls and limitations of environmental DNA surveys. *BMC Biol.* **2**, 13 (2004).
- Bass, D. *et al.* Polyubiquitin insertions and the phylogeny of Cercozoa and Rhizaria. *Protist* **156**, 149–161 (2005).
- Cavalier-Smith, T. & Chao, E. E. Phylogeny and classification of phylum Cercozoa (Protozoa). *Protist* **154**, 341–358 (2003).
- Titus, M. A. A class VII unconventional myosin is required for phagocytosis. *Curr. Biol.* **9**, 1297–1303 (1999).
- Tuxworth, R. I. *et al.* A role for myosin VII in dynamic cell adhesion. *Curr. Biol.* **11**, 318–329 (2001).
- Marchler-Bauer, A. *et al.* CDD: a curated Entrez database of conserved domain alignments. *Nucleic Acids Res.* **31**, 383–387 (2003).
- Bateman, A. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **32**, D138–D141 (2004).
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**, 4876–4882 (1997).
- Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574 (2003).
- Schmidt, H. A., Strimmer, K., Vingron, M. & von Haeseler, A. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18**, 502–504 (2002).
- Felsenstein, J. *Phylogeny* (Department of Genetics, University of Washington, Seattle, 1995).
- Minotto, L., Edwards, M. R. & Bagnara, A. S. *Trichomonas vaginalis*: Characterization, expression, and phylogenetic analysis of a carbamate kinase gene sequence. *Exp. Parasitol.* **95**, 54–62 (2000).
- Baldauf, S. L. & Palmer, J. D. Animals and fungi are each other’s closest relatives: congruent evidence from multiple proteins. *Proc. Natl Acad. Sci. USA* **90**, 11558–11562 (1993).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Preliminary sequence data were obtained from The Institute for Genomic Research website (<http://www.tigr.org>) and the Department of Energy Joint Genome Institute (JGI) website (<http://www.jgi.doe.gov>). We thank TIGR and DOE JGI for making data publicly available, A. A. Davies for comments and assistance with data management, and D. Soanes for PSI BLAST assistance. T.A.R. was supported by a BBSRC studentship. T.C.-S. thanks NERC for research grants and NERC and the Canadian Institute for Advanced Research Evolutionary Biology Program for Fellowship support.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.A.R. (thomr@nhm.ac.uk).

Architecture of floral branch systems in maize and related grasses

Erik Vollbrecht¹†, Patricia S. Springer¹†, Lindee Goh¹†, Edward S. Buckler IV² & Robert Martienssen¹

The external appearance of flowering plants is determined to a large extent by the forms of flower-bearing branch systems, known as inflorescences, and their position in the overall structure of the plant. Branches and branching patterns are produced by tissues called shoot apical meristems. Thus, inflorescence architecture reflects meristem number, arrangement and activity, and the duration of meristem activity correlates with branch length. The inflorescences of maize, unlike those of related grasses such as rice and sorghum, predominantly lack long branches, giving rise to the tassel and familiar corncob. Here we report the isolation of the maize *ramosa1* gene and show that it controls inflorescence architecture. Through its expression in a boundary domain near the nascent meristem base, *ramosa1* imposes short branch identity as branch meristems are initiated. A second gene, *ramosa2*, acts through *ramosa1* by regulating *ramosa1* gene expression levels. *ramosa1* encodes a transcription factor that appears to be absent in rice, is heterochronically expressed in sorghum, and may have played an important role in maize domestication and grass evolution.

Inflorescence architecture comprises the stereotypical number and arrangement of floral branches that characterizes each species of flowering plant^{1,2}. The presence or absence of long branches, for example, dictates the capacity for flower production. Therefore, in grasses including the domesticated cereals, branch length greatly affects grain-bearing capacity. Grass flowers are always produced on a specialized short branch called the spikelet, so that architecture in grasses is largely determined by iterations of branching before spikelet production, by whether or not branch primordia grow out, and by axis orientation in space^{3,4}. Branch length is central because of its macroscopic ramifications. The head of wheat, for instance, bears spikelets directly on the principal axis and thus appears quite different from a long branched inflorescence like that of rice.

Maize possesses a canonical grass architecture and two types of inflorescences, the tassel and the ear (Fig. 1), which differ in the presence of long, indeterminate (that is, bearing a large number of parts) branches at the tassel base^{5,6}. Otherwise, the ear and the upper tassel develop from morphologically similar spikes. The spike primordia bear many short, determinate (that is, having a fixed number of parts) branches; each short branch is called a spikelet pair because it bears two spikelets. The classical mutants *ramosa1* (*ra1*) and *ramosa2* (*ra2*) have long instead of short inflorescence branches. *ramosa1* was discovered in a farmer's field almost a century ago⁷ and has a branched, conical inflorescence (Fig. 1b, e) like that of many other grasses. Unlike other inflorescence mutants⁸, *ramosa1* is fully fertile and was initially classified as a new species, *Zea ramosa*^{7,9}, although it corresponds to a recessive mutation at a single locus on chromosome 7 (ref. 10). We isolated the *ra1* gene from maize and its close relatives, and from species in the sugar cane tribe, for which the last common ancestor with maize lived some 16 million years ago. Our findings address the genetic origins of the maize ear, suggest a general role for the *ramosa* genes in long-branch architecture in cereals, and implicate the *ramosa* pathway in the evolutionary diversification of grass inflorescence development.

ramosa1 encodes a transcription factor

During maize inflorescence development, the apex initiates successive primordia that mature as they are displaced towards the base¹¹. Thus, a chronology of developmental stages can be observed with a scanning electron microscope (SEM), from youngest at the apex to oldest at the base. In normal tassels, the primary inflorescence meristem initiated a few second-order meristems that were indeterminate (green arrowheads in Fig. 1h, i), and then switched to producing only determinate second-order meristems (red arrowheads in Fig. 1h, i). All second-order meristems in ears were determinate, producing paired spikelets (Fig. 1j) and straight, paired rows of kernels (Fig. 1d). In *ra1-R* mutants, no extra branch initiation was observed, but second-order meristems continued to grow into protruding second-order axes (Fig. 1m, n) that developed directly into the extra, long and indeterminate branches of mature mutant ears and tassels (Fig. 1o, p). Thus, the *ra1* gene imposes a determinate fate on second-order meristems.

To clone *ra1*, we used *Suppressor-mutator* (*Spm*) transposable elements in a directed transposon-tagging strategy¹². Three recessive alleles were recovered with somatically unstable (that is, mutable) inflorescence phenotypes (Fig. 1c, f), a hallmark of transposon-induced mutations. Genetic experiments with two alleles, *ra1-m2* and *ra1-m3*, identified tight linkage between the mutable phenotype and an autonomous *Spm* transposon (see Supplementary Information). Molecularly, these were independent *Spm* insertions that were not present in their progenitor chromosome (for example, Fig. 2a). We cloned DNA adjacent to *Spm* on the *ra1-m2* chromosome and used it to discover that the *ra1-m3* chromosome contained a different *Spm*, inserted 188-bp away. Plants with sectored tassels were mosaic for the *Spm* insertion, and the element was absent from *ra1-m* loci that had changed to stably mutant alleles (Fig. 2a). Two other alleles, *ra1-m1* and *ra1-m4*, also contained *Spm* insertions, all within a 690-bp window, identifying the *ra1* locus.

Plants with mutable phenotypes were chimaeras for *Spm* insertions,

¹Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA. ²USDA, ARS and Department of Plant Breeding, Cornell University, Ithaca, New York 14850, USA.

†Present addresses: Department of Genetics, Development and Cell Biology, Iowa State University, Ames, Iowa 50011, USA (E.V.). Department of Botany and Plant Science, Center for Plant Cell Biology, University of California, Riverside, California 92521, USA (P.S.). The Boston Consulting Group, 1 Exchange Place, Boston, Massachusetts 02109, USA (L.G.).

which was useful for inferring some aspects of *ra1* function. Clonal tassel sectors were scored as revertant (phenotypically normal) if a row of determinate second-order meristems was flanked by a row(s) of indeterminate second-order meristems. Pollen is derived from the inner cell layer (L2) in maize¹³, so we used pollen from individual flowers within revertant sectors to compare the genotypes of inner tissue layers with sector phenotype. In one example, pollen from only one of four narrow sectors (between four and ten spikelets long) transmitted the revertant allele. So the other three sectors were phenotypically normal even though their L2 cells were mutant. Therefore, transposon excision in other cells, such as the outer L1 layer, provided *ra1* activity that conferred determinacy. Similarly, large revertant sectors in the ear imposed determinacy on genetically mutant second-order meristems at sector boundaries (Fig. 1f). Thus, *ra1* functions cell-nonautonomously.

The transposon insertions interrupted a gene encoding a Cys₂-His₂

zinc-finger protein belonging to the plant-specific EPF subclass¹⁴. The EPF zinc-finger binds DNA via a short α -helix containing the amino acid sequence QALGGH¹⁴, which is conserved invariantly in the 28 and 33 single-finger EPF genes of *Arabidopsis* and rice, respectively. The *ra1* gene in maize and its orthologues in other grasses encode the variant QGLGGH (with a helix-relaxing Gly residue instead of Ala), suggesting that the RA1 zinc-finger may have unique functional attributes (Fig. 2d). Neither *Arabidopsis* nor rice appear to contain an orthologous gene (see Methods), but the most similar *Arabidopsis* EPF gene is *SUPERMAN*, which represses supernumerary stamens and is expressed in a boundary domain of floral meristems¹⁵. Strong mutant alleles *ra1-TN* and *ra1-R* contained point mutations in absolutely conserved residues in the zinc-finger (C51Y and H64N, respectively) and are probably null alleles¹⁵; weak alleles contained altered amino- and carboxy-terminal amino acid sequences (Fig. 2c). *ra1* RNA was detected only in developing inflorescences. The strong point mutant *ra1-R* and two weak alleles showed unaltered RNA

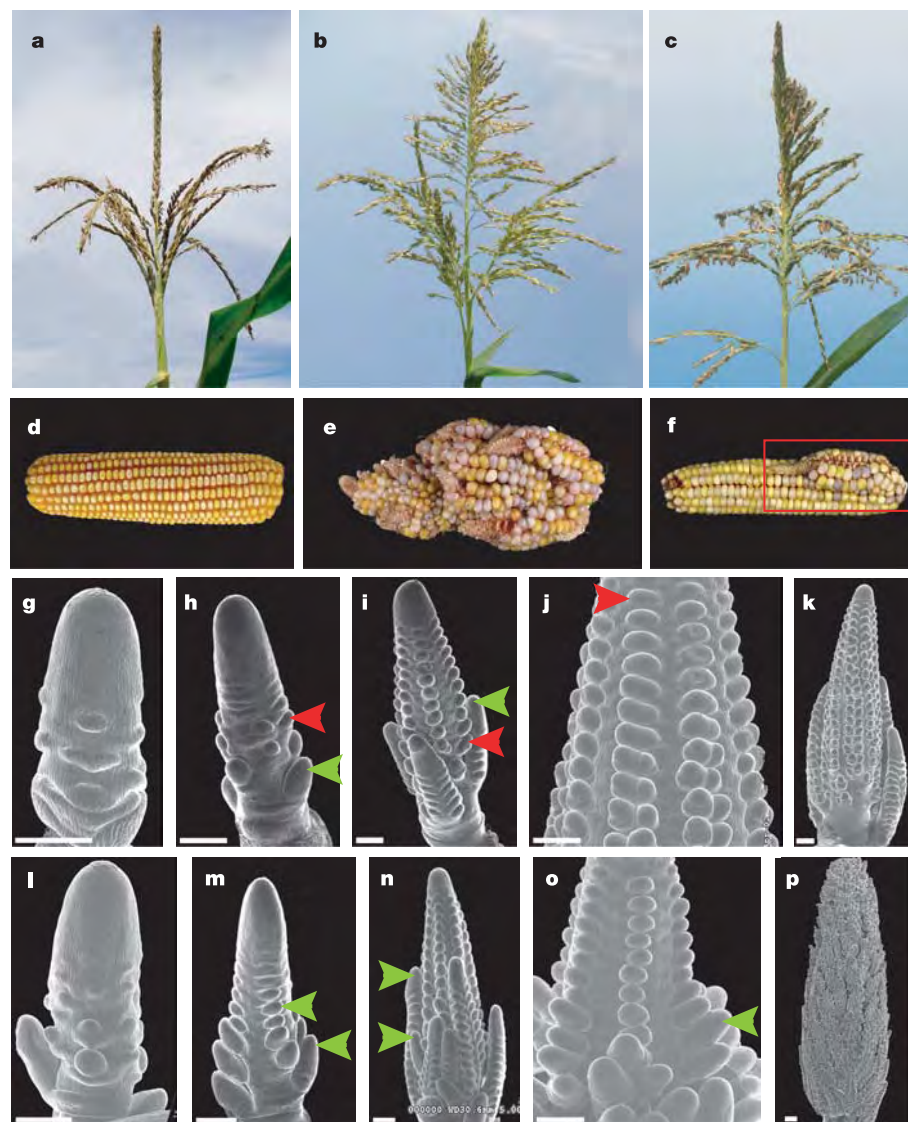


Figure 1 | Maize inflorescences. **a, d**, Normal tassel and ear. **b, e**, Mutant, highly branched *ra1-R* tassel and ear. **c, f**, Mutant *ra1-m2* tassel and ear are mosaics of normal and mutant tissue. The mutant ear sector (red box) contains spotted kernels and therefore *Spm* transposon activity. The zone of spotted kernels extends beyond the sector boundary into the normal portion. **g–k**, Scanning electron micrographs of tassel (**g–i, k**) and ear (**j**)

development in standard inbred B73 maize. Most second-order meristems in the tassel (for example, red arrowhead in **h**) and all secondary meristems in the ear (**j**) produce short, compact branches. **l–p**, Developing *ra1-R* tassel (**l–n, p**) and ear (**o**). Most second-order meristems in the tassel (green arrowheads in **m**) and in the ear (**o**) produce transformed, long branches. Scale bars, 250 μ m.

levels, suggesting that *ra1* transcriptional activation is not autoregulated (Fig. 2b).

ramosa1 expression predicts branch determinacy

ra1 expression was not detected during initiation or early growth of indeterminate second-order meristems (long branches), but only as determinate second-order (spikelet pair) meristems were initiated higher up the tassel (Fig. 3). *ra1* RNA was expressed at the junction between each determinate meristem and its parent, indeterminate axis, defining an adaxial meristem domain (Fig. 3a). As second-order meristems then branched to initiate third-order meristems, expression persisted across the base of the short secondary branch, strongest between the third-order branches (Fig. 3b), before ceasing. The only other developmental stage in which *ra1* was detected was

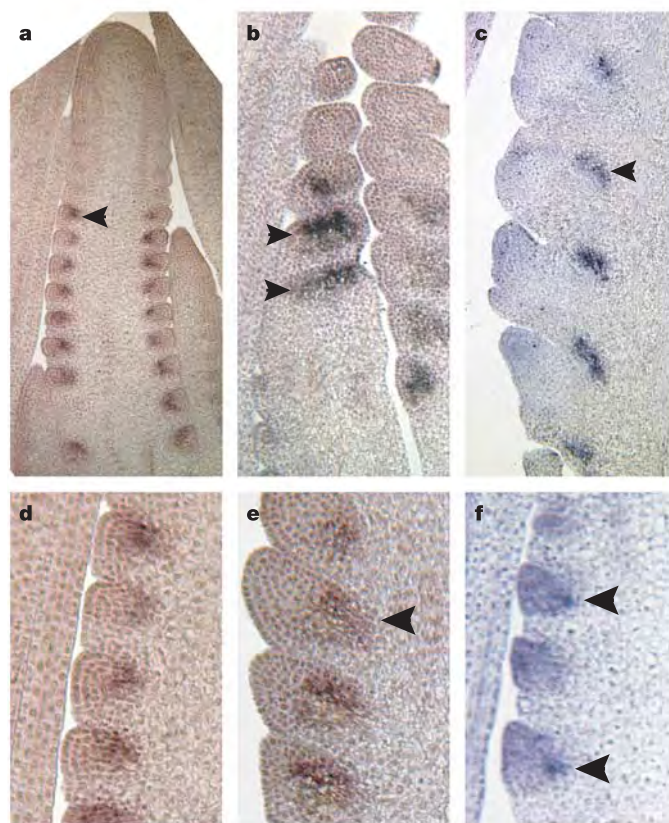
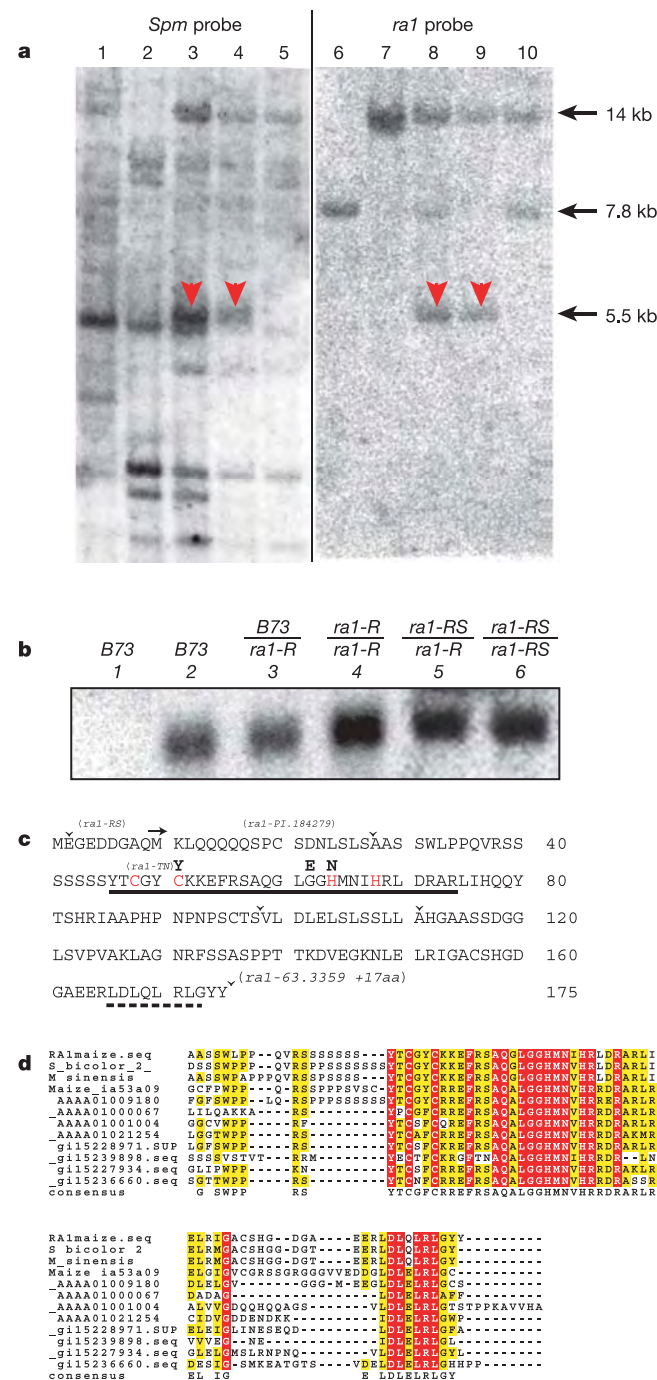


Figure 3 | *ra1* gene expression pattern. **a**, Longitudinal section of immature B73 tassel showing *ra1* expression as dark patches (for example, arrowhead) in recently initiated, second-order meristems. **b**, Immature tassel branch, sectioned parallel to the spikelet pair surface shows *ra1* expression across the broad base of the compact branch (arrowheads). **c**, Expression on the abaxial base of spikelet (third-order) meristems in the upper tassels. **d**, B73 (determinate second-order meristems). **e**, *ra1-R* (indeterminate second-order meristems), in which the basal expression domain is separate from the second-order meristem that has grown away from the primary axis. **f**, *ra2-R* (indeterminate second-order meristems), in which *ra1* is expressed in a properly located but greatly restricted domain.

Figure 2 | Transposon tagging, expression and sequence of the *ra1* gene.

a, DNA gel blot showing co-segregation of *ra1-m2* and *Spm*. Lanes 1–5, *Spm* probe; lanes 6–10, same blot probed with DNA flanking the *Spm* transposon that co-segregated with *ra1-m2*. Phenotypes and genotypes of lanes: 1 and 6, normal homozygous progenitor *ra1*⁺; lanes 2 and 7, mutant homozygous *ra1-R*; lanes 3, 4, 8 and 9, somatic mosaic of mutant and normal heterozygous *ra1-R/ra1-m2*; lanes 5 and 10, normal, heterozygous *ra1-R/ra1-m2rev4* (a stable, normal allele derived from *ra1-m2*). The 5.5-kb fragment (arrowheads) contains both *Spm* and *ra1*. The 7.8-kb progenitor *ra1*⁺ fragment is restored by *Spm* excision, either in variable stoichiometries when excision occurs somatically (lanes 8 and 9) or entirely when it occurs germinally (lane 10). The 14-kb fragment (lanes 7–10) derives from *ra1-R*. **b**, RNA gel blot. Lane 1, vegetative shoot apices; lanes 2–6, equally staged immature tassels. **c**, Single letter amino-acid code translation of the *ra1* open reading frame. Solid line indicates the zinc-finger, dotted line highlights the EAR domain. Changes in mutant alleles are shown above the sequence, adjacent to the allele designation: insertion locations indicated by carets; *ra1-RS* alternative methionine codon indicated with an arrow; amino acid changes shown. **d**, Multiple amino acid sequence alignment of the zinc-finger and EAR domains of RA1 from maize (RA1maize), sorghum (*S_bicolor_2*) and *Miscanthus* (*M_sinensis*), and of EPFs from maize, rice (AAAA...) and *Arabidopsis* (*gi...*). Absolutely conserved residues highlighted in red, highly conserved residues in yellow.

slightly later, when *ra1* was expressed on the lower (abaxial) edge of the differentiating third-order branch (spikelet) (Fig. 3c).

The early *ra1* expression domain suggested either that *ra1* marks the persistent second-order meristem until it initiates third-order meristems or that *ra1* transiently marks a boundary between indeterminate and determinate branch axes. In the normal inflorescence, the spikelet pair (second-order) meristem remains close to the primary axis, making it difficult to distinguish between these hypotheses (Fig. 3d). We therefore examined *ra1* expression in strong *ra1-R* mutants, in which axis growth carries second-order meristems away from the primary axis. As second-order meristems initiated, the *ra1-R* transcript was detected in the normal, adaxial domain. After the indeterminate second-order meristem was borne away from the axis, expression persisted at the junction between second-order meristems and the primary axis (Fig. 3e). Expression in *ra2-R* mutants was reduced (Fig. 3f), as discussed below. Thus, *ra1* imposes determinacy on nascent second-order meristems, not in the meristem *per se*, but in a boundary domain near the junction with the indeterminate primary axis.

Traversing from base to apex, the main axis of wild-type tassels (Fig. 4a) bore first long and then strictly short spikelet pair branches (Fig. 4k, bottom and top panels). Strong *ra1-R* mutants (Fig. 4d) first produced a few extra long branches bearing spikelet pairs (Fig. 4j, green), followed by several transformed, mixed-fate branches bearing both spikelet pairs and single spikelets (Fig. 4j, yellow). Proceeding apically, transformed branches bore multiple, single spikelets (Fig. 4j, orange). Successive spikelet multimers bore progressively fewer spikelets until they occurred strictly in pairs, defining a short, apical spike. Normal tassels produced simple, long branches at their base (Fig. 4a), but strong mutants produced compound, branched side axes (Fig. 4d) bearing a similar range of transformed branch fates as the main axis. Thus, second-order meristems that lacked *ra1* boundary function adopted a relatively continuous range of fates, revealing an underlying apical–basal gradient of indeterminacy.

Weak *ra1-RS* mutants (Fig. 4c) produced the same range and relative arrangement of altered second-order branch fates as strong mutants, but transition types occurred across a smaller region, resulting in a longer apical spike (Fig. 4j). This constricted transition region indicates a threshold-dependent interaction, in which second-order branch determinacy is controlled by the interplay of *ra1* boundary activity (allele strength) and position in an apical–basal gradient.

A genetic pathway regulates inflorescence branching

ramosa2 (*ra2*) and *liguleless2* (*lg2*) mutants have more (Fig. 4b) or fewer (Fig. 4e) long branches in the inflorescence, respectively^{16,17}. Using SEM, we observed that in *ra2-R* mutants, extra long branches developed directly from transformed, indeterminate second-order meristems (not shown), and resembled those in weak *ra1* mutants (Fig. 4j). Inbred B73 maize produces a compact ear with straight rows of kernels (Fig. 4l). In B73, *ra2-R* and weak *ra1-RS* mutants each produced fertile ears with crooked rows owing to slight second-order branch indeterminacy (Fig. 4m, n), whereas strong *ra1-R* mutants produced highly branched, functionally sterile ears (Fig. 4o). Notably, *ra2-R*;*ra1-RS* double mutants produced highly branched ears (Fig. 4p) that resembled those of *ra1-R* (strong) mutants. We therefore examined *ra1* expression in *ra2-R* mutants. RNA gel blots indicated that it was considerably reduced (data not shown), consistent with RNA *in situ* hybridizations in which *ra1* was expressed in its normal position at the base of second-order meristems, but in a highly constricted domain that appeared as a small speck (Fig. 3f). Thus, *ra2* regulates accumulation of *ra1* transcripts, placing the two genes in a single (*ramosa*) genetic pathway, with *ra2* upstream of *ra1*.

In *lg2-R* mutants, long branches at the base of the tassel failed to initiate and were replaced by bare nodes¹⁷ (Fig. 4e, g, j). Thus, *lg2* function is required for either identity or outgrowth of these long branches. To distinguish between these possibilities, we constructed double mutants of *lg2-R* and each strong *ramosa* mutation, *ra1-R* and

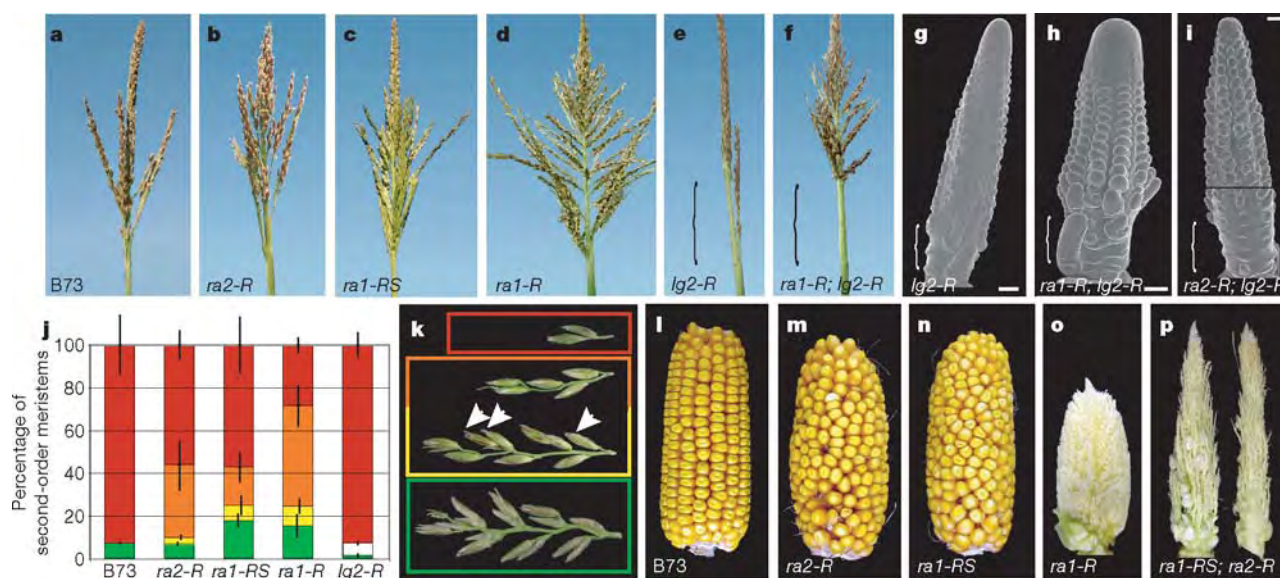


Figure 4 | Developmental genetics of long branch pathways. All alleles are in the maize inbred B73 (a, l) genetic background. b–d, *ra2-R* (b), weak *ra1-RS* (c) and strong *ra1-R* (d) mutant phenotypes. e, *lg2-R* mutants form bare nodes instead of basal branches (bracketed region). f, Additive phenotype of *ra1-R*;*lg2-R* double mutant separates *lg2* and *ra1* pathways. g–i, Scanning electron micrographs of immature tassels. Bare nodes of all mutants containing *lg2-R* have the same developmental basis: failure to initiate or maintain second-order meristems. j, k, Apical–basal distribution of tassel second-order meristems that developed as canonical, determinate spikelet

pairs (red; top panel in k), slightly indeterminate ‘spikelet multimer’ branches (orange; upper middle panel in k), mixed, indeterminate long branches bearing single (arrow) and paired (double arrow) spikelets (yellow; lower middle panel in k), and canonical, indeterminate long branches that bear spikelet pairs (green; bottom panel in k). White box indicates bare nodes. Error bars indicate standard deviation of the class below. l–p, Mature ears. *ra2-R* (m) and *ra1-RS* (n) have similar, weak phenotypes. The double mutant ear (p) is highly branched, like *ra1-R* (o), implying that *ra1* and *ra2* affect the same determinacy process. Scale bars, 250 μm.

ra2-R. Both double mutants produced tassels with bare nodes basally, and transformed, long branches apically (Fig. 4f, h, i). Thus, *lg2* is specifically required for branch outgrowth only at the base of the tassel, defining a distinct genetic pathway that regulates long branching upon the transition to flowering¹⁷.

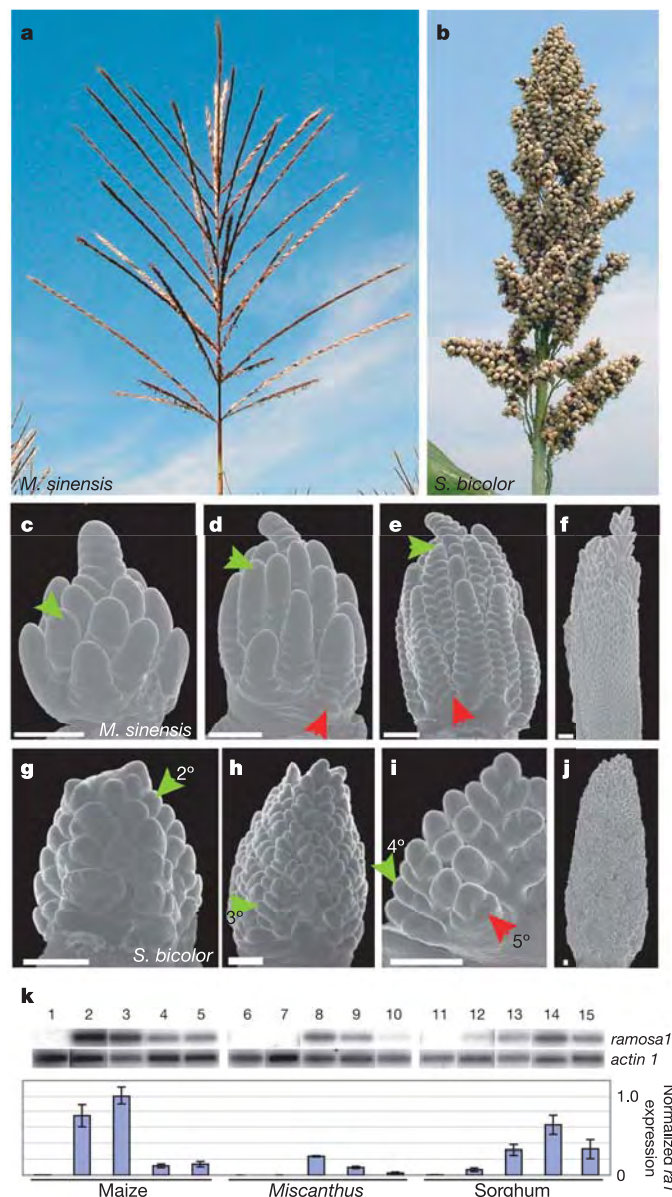


Figure 5 | Comparative development and *ra1* expression in Panicoid grasses. **a, b**, Mature inflorescences. *Miscanthus sinensis* (**a**) is moderately branched like maize, *Sorghum bicolor* (**b**) is reiteratively branched like maize *ramosa* mutants. **c–f**, Scanning electron micrographs of *M. sinensis* development. All second-order meristems (2°) are indeterminate (**c, d**, green arrows) and all third-order meristems (3°) are determinate (**d, e**, red arrows) and produce spikelet pairs. **g–j**, *S. bicolor* development. Multiple orders of indeterminate branching occur (**g–i**, green arrows) before fifth-order meristems (**i**, red arrow) typically become spikelet pairs. **k**, Semi-quantitative RT–PCR of *ra1* gene expression, showing raw data (above) and *ra1* normalized to *actin1* levels (below). RNA was isolated from vegetative apices (lanes 1, 6 and 11), carefully staged tassels of B73 (lanes 2–5), and analogous stages of *Miscanthus* (7–10) and sorghum (12–15) that match **c–f** and **g–j**, respectively. The abrupt transition from indeterminate to determinate second-order meristems in maize and *Miscanthus* is accompanied by an abrupt onset of *ra1* gene expression. In sorghum, delayed production of spikelet pairs correlates with a protracted onset of *ra1* expression. Error bars indicate s.d. Scale bars, 250 μ m.

The *ra1* pathway in grass evolution and domestication

Although maize normally produces spikelets only in pairs, mutants with different levels of *ra1* activity produce long branches and spikelet multimers, resembling architectures of other grasses¹⁸. The popular ornamental grass *Miscanthus sinensis* produces a visually simple inflorescence with discrete, long branches (Fig. 5a), similar to the base of the maize tassel. The crop plant *Sorghum bicolor*, on the other hand, generates a dense, multi-branched head (Fig. 5b) that resembles a *ramosa* mutant. We isolated the *ra1* orthologue from each species (Fig. 2d). In maize, *ra1* expression was highest when second-order meristems were initiated, imposing determinacy on these spikelet pairs. Expression dropped off suddenly as third- and higher-order meristems were produced on spikelet pair axes (Fig. 5k, lanes 1–5). During *Miscanthus* development, all second-order meristems were indeterminate (green arrows in Fig. 5c–e). This extended phase of indeterminate second-order meristems correlated with a delayed, similarly narrow window of *ra1* expression (Fig. 5k, lanes 6–10), which was highest when strictly determinate third-order meristems became spikelet pairs (red arrows in Fig. 5d, e) and then dropped off as fourth- and higher-order meristems were produced on spikelet pair axes. Thus, maize and *Miscanthus* showed similar expression dynamics, reflecting similar architectures in which spikelet pairs are all produced from a single order of meristem. Expression dynamics differed for sorghum, in which all second-, most third- and even fourth-order meristems were indeterminate (Fig. 5g–i), such that spikelet pairs were derived from third- to fifth-order meristems. The late acquisition of spikelet pair fate in sorghum, together with the continual indeterminate branching and extended phase of *ra1* expression (Fig. 5k, lanes 11–15), suggest that *ra1* activity reaches a threshold level sufficient to impose determinate spikelet pair fate at a relatively late developmental stage. Thus, *ra1* activity regulates long branch architecture similarly in these three species, by imposing spikelet pair identity on the appropriate order of meristem.

A gene that regulates natural inflorescence diversity in the grasses might be a target for selection during the domestication of a crop species like maize, in which the ear inflorescence has undergone such intense selective pressure as to be regarded a monstrosity relative to typical wild species¹⁹. We surveyed the sequence of the *ra1* locus in a panel of diverse, inbred strains of maize, which had previously been examined for multiple loci that have or have not undergone selection²⁰ (see Supplementary Information). Silent site diversity at *ra1* ($\pi = 0.0010$) was >10-fold less than that of typical, non-selected maize genes ($\pi = 0.0107$), but >2-fold lower than observed for starch pathway genes ($\pi = 0.0027$ – 0.0050), which were targets of human selection during maize domestication and improvement²⁰. We tested for selection at *ra1* using Hudson–Kreitman–Aguadé (HKA) tests²¹, which were highly significant ($P < 0.0001$). These data all indicate that *ra1* was a target of positive directional selection in the maize lineage. Consistent with selection, the diverse inbreds contained only three haplotypes at *ra1*, and we detected no evidence for recombination. Such low recombination is uncommon in maize^{22–24}.

Discussion

The plant morphologist W. Troll surveyed some 40,000 species to infer that the flowering plant shoot is comprised of zones defined by branch determinacy and their relationship to the parent shoot²⁵ (Fig. 6): the E zone (final inflorescence), the subapical ZE (zone of inflorescence enrichment), the INZ (zone of inhibition) and the basal IZ (innovation zone of renewal growth). Conspicuous in the transition from vegetative to inflorescence shoot is the precocious outgrowth of lateral meristems, so that the ZE usually contains many side branches (paracladia) and the INZ relatively few. For all zones, paracladia repeat the morphology of the principal axis at their site of attachment—the lower the paracladium, the more extensive is the repetition. Thus, Troll's model considers the inflorescence in

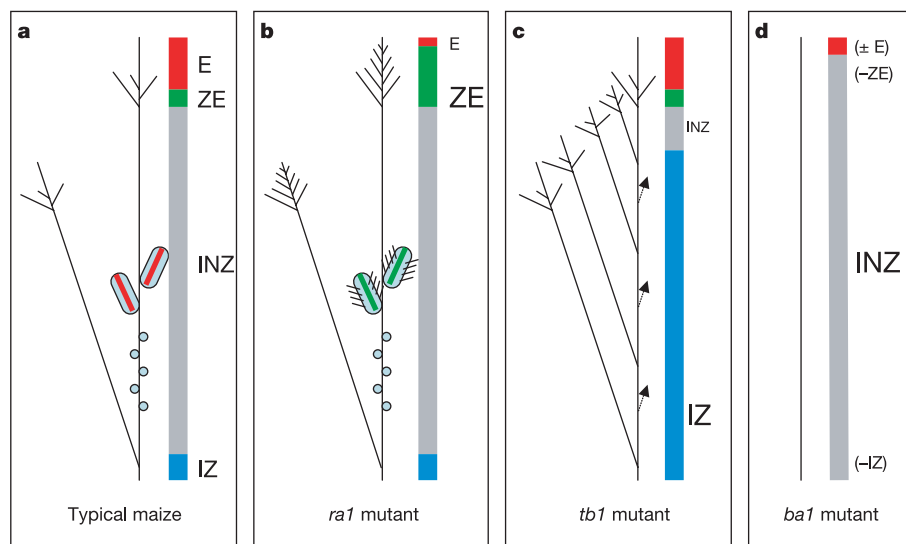


Figure 6 | A model for heterochronic modulation of inflorescence and plant architecture. **a**, Normal maize shoot (schematic) illustrates elements of Troll's holistic model (coloured bars on right): the shoot zones include the zone of innovation (IZ, blue bar), the zone of inhibition (INZ, grey bar) with ear shoots mostly unelaborated (small circles) and a few elaborated shoots (ovals), the zone of enrichment (ZE, green bar) and the final inflorescence (E zone, red bar). Paracladia in IZ and ZE mimic their parent shoot above the

point of origin. **b–d**, For mutants, only affected zones are labelled; altered font size reflects enlarged or contracted zones. **b**, *ra1*-*R* mutants expand the ZE at the expense of the E zone. **c**, *tb1* mutants expand the IZ into the INZ (only one half of the *tb1* shoot is shown). **d**, Depending on the allele strength, *ba1* mutants expand the upper INZ to eliminate the IZ and ZE entirely, and decrease the length of the E zone.

the context of the whole shoot system. In maize, for example, tillers in the basal IZ reproduce the entire plant, but higher lateral branches are short and tipped by an inflorescence (the ear) that mimics just the E zone of the tassel. Except for basal long branches in the tassel, the majority of the ZE is suppressed by *ra1* so that a 'determinacy gradient' is only evident in the absence of *ra1* function. In the ear branch, this suppression affords the remarkable packing of kernels in the corn cob.

Each of the zones envisioned by Troll arises in a fixed temporal sequence, with later structures at the apex and earlier ones at the base. Thus, genes such as *ra1* that suppress elaboration of a given zone may be considered heterochronic: *ra1* accelerates later stages of branch development (spikelet pairs) at the expense of earlier ones (long branches). The gradation of mutant branch types, and the shift in this gradation with allele strength, mimics altered temporal transitions of other heterochronic mutations in maize²⁶. Changes in the developmental timing of *ra1* expression, attributable for example to promoter rearrangements at *ra1* or to alterations in *ra2*, underlie a difference in inflorescence architecture in certain maize alleles as well as in related grasses. A similar model can be proposed for modulation of earlier zones by teosinte *branched1* (*tb1*) and *barren stalk1* (*ba1*), which regulate multiple lateral meristems in vegetative and reproductive shoots, respectively, and modulation of later zones by *indeterminate spikelet1* (*ids1*) and *branched silkleless1* (*bd1*)^{8,27–29}. *lg2* has also been described as heterochronic¹⁷. Interestingly, these genes all encode transcription factors. Moreover, *ra1*, *ba1* and *bd1* are all expressed in boundary domains adjacent to the meristems they regulate, and *ra1* (at least) displays cell-nonautonomy. The products of these genes may affect zones by regulating a mobile signal for meristem determinacy, or be mobile themselves. Other genes affect inflorescence architecture without modifying temporal zones, for example by regulating meristem size (for example, *fasciated ear2*, *thick tassel dwarf1* and *knotted1*^{30–32}).

Almost a century ago, the similarity of *ra1* mutants to other grasses led to the proposal that *ra1* represented a 'revertant' or atavistic evolutionary form³³. Our data indicate directional selection for a narrow set of *ra1* alleles, either during maize domestication from its wild ancestor teosinte or during agricultural improvement in the

subsequent 9,000 yr. We propose that selection was most likely for suppressed branching in the ear. Modern inbreds never have long-branched cobs, yet they are diverse for tassel branch number. Furthermore, although the tiny, two-rowed teosinte ear is also unbranched, it is highly divergent from its massive, modern maize counterpart. Ear improvement probably involved selecting genes that increased the size of primary meristems³⁰, which in turn produced more second-order meristems and may have intensified a requirement to keep second-order meristems determinate. The selected trait is probably also still variable in extant teosinte, so we predict that teosinte *ra1* alleles will be more diverse molecularly, and perhaps even phenotypically, when introduced into maize. Key morphological differences between maize and teosinte can be accounted for by genetic changes in five major chromosomal regions; these include *tb1* and possibly *ba1*^{27,34,35}, which are highly conserved in sequence and function across most of the grass phylogeny^{27,36}. Such purifying selection is expected to constrain the ability of *tb1* and *ba1* to confer evolutionary novelty. However, domestication was a complex process that acted on an enormous number of traits and their controlling genes³⁷. *ra1* is not within the five regions, and regulates a specific meristem determinacy switch⁹, the implementation of which varies even between relatively closely related species. Moreover, *ra1* sequence diverges rapidly among these species, and the gene appears to be absent in distantly related rice, for which spikelet-bearing axes are indeterminate. Thus, *ra1* may function in the evolution of developmental diversity^{38,39} and be useful for molecular breeding and crop development.

METHODS

Maize genetics and *ra1* gene cloning. For transposon tagging, *o2*^{m20}::*Spm* ears were fertilized with *ra1-R o2 gl1* pollen and the progeny were screened at maturity for branched tassels and ears. Three mutable alleles, *ra1-m1*, *ra1-m2* and *ra1-m3*, were identified among approximately 50,000 plants. *ra1-m4* was isolated from 20,000 plants from a cross between *bz1-mum9* ears and *ra1-R o2 gl1;sh1 bz1 wx1* pollen. All four alleles were caused by insertion of the *Spm* transposable element (see Supplementary Information), and a co-segregating DNA fragment from the *ra1-m2* allele was isolated by size-selecting *EcoRI*-digested DNA in an agarose gel and ligating into a λ -ZAP vector (Stratagene). Approximately 50,000 plaques were screened by hybridization with an *Spm*

probe. The DNA insert in one positive clone was isolated and sequenced. Transcript ends were determined by rapid amplification of cDNA ends (RACE)⁴⁰ using total RNA isolated from young inflorescences with Trizol (Gibco). DNA from various inbred and mutant *ra1* alleles was amplified by polymerase chain reaction (PCR) using gene-specific primers (and with transposon-specific primers for *Spm* alleles, as guided by DNA gel blots) and then sequenced. RNA gel blots used 5 µg total RNA per sample; nucleic acid gel blots were performed as described³². Maize alleles used for expression analysis were first crossed between three and six times to inbred B73. Maize and *Sorghum bicolor* (cultivar Btx623) were grown outdoors or in a greenhouse. Seasonally harvested *Miscanthus sinensis* was grown on the Cold Spring Harbor Laboratory grounds. To isolate *ra1* from sorghum, we screened a bacterial artificial chromosome (BAC) library (CUGI) with a maize probe and subcloned and sequenced a 6-kb *XbaI* restriction fragment that contained the tandemly duplicated locus. The same region was PCR-amplified and directly sequenced from sorghum genomic DNA for verification. *ra1* was isolated from *Miscanthus* by PCR (primers RA_35 AACGACGGATCACGCTCTGTGTGTC, and Sb_1826_AS TGGACTCTGTGTCGTTGTTGGA) and TA-cloned, and multiple clones were sequenced until an identical sequence was obtained from different clones. EPF genes in rice and *Arabidopsis* were identified by searching proteome databases (MATDB, <http://mips.gsf.de/proj/thal/db/> and Gramene, <http://www.gramene.org>) with blastp using the Cys₂-His₂ region of RA1 and a cutoff *p*-value of 1, followed by manual elimination of false positives. Lack of a *ra1* orthologue in rice or *Arabidopsis* was concluded based on not finding a conserved protein with the QGLGGH motif or with sequence similarity other than in the highly conserved, short Cys₂-His₂ and EAR motifs.

Phenotype and RNA expression analysis. All developmental, morphological and double mutant phenotype analyses of maize were on material converged for at least four generations into the inbred B73. For morphometrics (Fig. 4j), 6–8 mature tassels of each genotype were dissected, classifying every second-order branch. The number of second-order meristems in each class was averaged and normalized against the total average number of second-order meristems. The bare nodes of *lg2* mutant tassels may be undercounted, as they are difficult to score unambiguously other than by SEM. For SEM of all species, freshly dissected inflorescences were mounted on stubs in silver paint (Electron Microscopy Sciences) and examined in an S-3500N scanning electron microscope (Hitachi) at high vacuum, using 3–10 kV accelerating voltage and a secondary electron detector. For maize RNA *in situ* hybridizations, inflorescence primordia were dissected, vacuum infiltrated and fixed at 4 °C overnight in 4% formaldehyde, prepared fresh from paraformaldehyde (Sigma) in phosphate-buffered saline. *In situ* hybridizations were performed as described⁴¹. Semi-quantitative PCR with reverse transcription at *ra1* was performed with primers RA_53 (GCCGCCAC AGGTAAGGTCG) and RA_49 (GCCAGTCTAAGCTGAAGAT CCA). RT-PCR with primers specific for each sorghum repeat determined that the upstream copy is hardly (but detectably) transcribed at all stages tested, but RA_53 did not amplify the upstream (untranscribed) copy at the annealing temperature used (63.5 °C). A 'grass actin' primer set was designed on the basis of actin expressed-sequence-tag (EST) sequences from maize, sorghum, barley, *Setaria italica* and rice (primers actin_F GTMARCAACTGGGAYGACATGGA GAA and actin_B ACRTCRCACTTCATGATRGAGTTGTABGT, where M, R, Y and B are standard degenerate nucleotide codes). PCR assays of 23, 25 and 27 cycles were processed by DNA gel blot and hybridization. A pair of cycles showing linear changes for both the actin gene and *ra1* was selected to quantify average expression levels using a Fuji phosphorimager.

Nucleotide diversity analysis. A 1.5-kb *ra1* fragment that includes the transcribed region plus ~400 bp upstream and downstream was isolated by PCR (primers RA_48 TCAACGTGGTCAAAGTTGTGTGTG and RA_36 CAAGGTG CACCCACAACATTGAC) and sequenced from a set of 30 maize inbred lines²⁰. Sequence alignments were made in ClustalW and adjusted by eye. Nucleotide diversity statistics were calculated using DnaSP⁴². HKA tests with neutral loci that were sequenced in the same panel of inbreds were performed by the direct method in DnaSP using *Tripsacum dactyloides* (PI 595898) as an outgroup (primers RA_28 CGTGGCTGATCTCAATCTCAA and RA_11 TGCACCTGC ACGTACCATTTGTAG), and scores were combined as previously described²⁰. PCR products from *Tripsacum* were TA-cloned and multiple clones sequenced until an identical sequence was obtained from different clones.

Received 9 March; accepted 6 June 2005.

Published online 24 July 2005.

- Weberling, F. *Morphology of Flowers and Inflorescences* (Cambridge Univ. Press, Cambridge, 1989).
- Sussex, I. M. & Kerk, N. M. The evolution of plant architecture. *Curr. Opin. Plant Biol.* **4**, 33–37 (2001).
- Clifford, H. in *Grass Systematics and Evolution* (eds Soderstrom, T., Hilu, K., Campbell, C. & Barkworth, M.) 21–30 (Smithsonian Institution Press, Washington DC, 1987).
- Doust, A. N. & Kellogg, E. A. Inflorescence diversification in the panicoid "bristle grass" clade (Paniceae, Poaceae): evidence from molecular phylogenies and developmental morphology. *Am. J. Bot.* **89**, 1203–1222 (2002).
- McSteen, P., Laudencia-Chinguanco, D. & Colasanti, J. A fleret by any other name: control of meristem identity in maize. *Trends Plant Sci.* **5**, 61–66 (2000).
- Veit, B., Schmidt, R. J., Hake, S. & Yanofsky, M. F. Maize floral development: new genes and old mutants. *Plant Cell* **5**, 1205–1215 (1993).
- Gernert, W. A new subspecies of *Zea mays* L. *Am. Nat.* **46**, 616–622 (1912).
- Chuck, G., Muszynski, M., Kellogg, E., Hake, S. & Schmidt, R. J. The control of spikelet meristem identity by the *branched silkless1* gene in maize. *Science* **298**, 1238–1241 (2002).
- Postlethwait, S. N. & Nelson, O. E. Characterization of development in maize through the use of mutants. I. The *polytypic* (*Pt*) and *ramosa-1* (*ra1*) mutants. *Am. J. Bot.* **51**, 238–243 (1964).
- Emerson, R., Beadle, G. & Fraser, A. A summary of linkage studies in maize. *Cornell Univ. Agric. Experiment Station Memoir* **180**, 3–83 (1935).
- Cheng, P. C., Greyson, R. I. & Walden, D. B. Organ initiation and the development of unisexual flowers in the tassel and ear of *Zea mays*. *Am. J. Bot.* **70**, 450–462 (1983).
- McClintock, B. Induction of instability at selected loci in maize. *Genetics* **38**, 579–599 (1953).
- Dawe, R. K. & Freeling, M. Clonal analysis of the cell lineages in the male flower of maize. *Dev. Biol.* **142**, 233–245 (1990).
- Takatsui, H. Zinc-finger proteins: the classical zinc finger emerges in contemporary plant science. *Plant Mol. Biol.* **39**, 1073–1078 (1999).
- Sakai, H., Medrano, L. J. & Meyerowitz, E. M. Role of *SUPERMAN* in maintaining *Arabidopsis* floral whorl boundaries. *Nature* **378**, 199–203 (1995).
- Nickerson, N. H. & Dale, E. E. Tassel modifications in *Zea mays*. *Ann. Mo. Bot. Gard.* **42**, 195–211 (1955).
- Walsh, J. & Freeling, M. The *liguleless2* gene of maize functions during the transition from the vegetative to the reproductive shoot. *Plant J.* **19**, 489–495 (1999).
- Kellogg, E. in *Grasses: Systematics and Evolution* (eds Jacobs, S. & Everett, J.) (CSIRO, Melbourne, 2000).
- Kellogg, E. A. Plant evolution: the dominance of maize. *Curr. Biol.* **7**, R411–R413 (1997).
- Whitt, S. R., Wilson, L. M., Tenaillon, M. I., Gaut, B. S. & Buckler, E. S. Genetic diversity and selection in the maize starch pathway. *Proc. Natl Acad. Sci. USA* **99**, 12959–12962 (2002).
- Hudson, R., Kreitman, M. & Aguade, M. A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153–159 (1987).
- Tenaillon, M. I. *et al.* Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Proc. Natl Acad. Sci. USA* **98**, 9161–9166 (2001).
- Remington, D. L. *et al.* Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc. Natl Acad. Sci. USA* **98**, 11479–11484 (2001).
- Wilson, L. M. *et al.* Dissection of maize kernel composition and starch production by candidate gene association. *Plant Cell* **16**, 2719–2733 (2004).
- Troll, W. *Die Infloreszenzen: Typologie und Stellung im Aufbau des Vegetationskörpers* (Fischer, Stuttgart, 1964).
- Poethig, R. Phase change and the regulation of shoot morphogenesis in plants. *Science* **250**, 923–930 (1990).
- Gallavotti, A. *et al.* The role of *barren stalk1* in the architecture of maize. *Nature* **432**, 630–635 (2004).
- Doebley, J., Stec, A. & Hubbard, L. The evolution of apical dominance in maize. *Nature* **386**, 485–488 (1997).
- Chuck, G., Meeley, R. & Hake, S. The control of maize spikelet meristem identity by the *APETALA-2*-like gene *indeterminate spikelet1*. *Genes Dev.* **12**, 1145–1154 (1998).
- Taguchi-Shiobara, F., Yuan, Z., Hake, S. & Jackson, D. The *fasciated ear2* gene encodes a leucine-rich repeat receptor-like protein that regulates shoot meristem proliferation in maize. *Genes Dev.* **15**, 2755–2766 (2001).
- Bommert, P. *et al.* *thick tassel dwarf1* encodes a putative maize ortholog of the *Arabidopsis* *CLAVATA1* leucine-rich repeat receptor-like kinase. *Development* **132**, 1235–1245 (2005).
- Vollbrecht, E., Reiser, L. & Hake, S. Shoot meristem size is dependent on inbred background and presence of the maize homeobox gene, *knotted1*. *Development* **127**, 3161–3172 (2000).
- Collins, G. Hybrids of *Zea tunicata* and *Zea ramosa*. *Proc. Natl Acad. Sci. USA* **3**, 345–349 (1917).
- Doebley, J. & Stec, A. Inheritance of the morphological differences between maize and teosinte: comparison of results for two F₂ populations. *Genetics* **134**, 559–570 (1993).
- Wang, R. L., Stec, A., Hey, J., Lukens, L. & Doebley, J. The limits of selection during maize domestication. *Nature* **398**, 236–239 (1999).
- Takeda, T. *et al.* The *OsTBI* gene negatively regulates lateral branching in rice. *Plant J.* **33**, 513–520 (2003).

37. Doebley, J. The genetics of maize evolution. *Annu. Rev. Genet.* **38**, 37–59 (2004).
38. Kellogg, E. A. Evolution of developmental traits. *Curr. Opin. Plant Biol.* **7**, 92–98 (2004).
39. Martienssen, R. The origin of maize branches out. *Nature* **386**, 443–444 (1997).
40. Sambrook, J. & Russell, D. W. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 2001).
41. Jackson, D., Veit, B. & Hake, S. Expression of maize *KNOTTED1* related homeobox genes in the shoot apical meristem predicts patterns of morphogenesis in the vegetative shoot. *Development* **120**, 405–413 (1994).
42. Rozas, J. & Rozas, R. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**, 174–175 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. Mulligan for plant care, Z. Lippman and C. Kopec for help with *in situ* hybridization and SEM, R. J. Schmidt for producing and sharing the *ra1-RS* allele, D. Jackson for discussions and V. Irish for commenting on the manuscript. E.V. was a DOE-Energy Biosciences postdoctoral fellow of the Life Sciences Research Foundation. L.G. was supported by the Cold Spring Harbor Undergraduate Research Program. Grant support was provided by the Agricultural Research Service of the USDA (to E.S.B.), the National Research Initiative of the USDA CSREES (to R.M.), and by the NSF Plant Genome Research Program (to E.S.B. and R.M.).

Author Information DNA sequences reported here have been deposited in GenBank under accession numbers AY957396–AY957399 and DQ013174–DQ013203. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.M. (martiens@cshl.org).

Early planetesimal melting from an age of 4.5662 Gyr for differentiated meteorites

Joel Baker¹, Martin Bizzarro^{2,3}, Nadine Wittig^{2,3}, James Connelly⁴ & Henning Haack³

Long- and short-lived radioactive isotopes and their daughter products in meteorites are chronometers that can test models for Solar System formation^{1,2}. Differentiated meteorites come from parent bodies that were once molten and separated into metal cores and silicate mantles. Mineral ages for these meteorites, however, are typically younger than age constraints for planetesimal differentiation^{3–5}. Such young ages indicate that the energy required to melt their parent bodies could not have come from the most likely heat source⁶—radioactive decay of short-lived nuclides (²⁶Al and ⁶⁰Fe) injected from a nearby supernova—because these would have largely decayed by the time of melting. Here we report an age of 4.5662 ± 0.0001 billion years (based on Pb–Pb dating) for basaltic angrites, which is only 1 Myr younger than the currently accepted minimum age of the Solar System⁷ and corresponds to a time when ²⁶Al and ⁶⁰Fe decay could have triggered planetesimal melting. Small ²⁶Mg excesses in bulk angrite samples confirm that ²⁶Al decay contributed to the melting of their parent body. These results indicate that the accretion of differentiated planetesimals pre-dated that of undifferentiated planetesimals, and reveals the minimum Solar System age to be 4.5695 ± 0.0002 billion years.

We have reinvestigated the chronometry of basaltic angrites taking advantage of improved methods for determining absolute and relative ages (see the Supplementary Information). Angrites are differentiated meteorites with igneous textures and largely composed of Ca–Al–Ti-rich pyroxene, Ca-rich olivine and anorthitic plagioclase⁸. Initial Sr isotope data⁹ indicates that angrites are the most primitive basaltic meteorites. Earlier Pb–Pb dating of minerals from angrites Angra dos Reis (ADOR), LEW86010 and D'Orbigny yielded the most precise absolute ages for basaltic meteorites of 4.555–4.558 billion years (Gyr; refs 9, 10), underpinning the mapping of the short-lived ⁵³Mn–⁵³Cr (ref. 5) and, in part, ¹⁸²Hf–¹⁸²W (ref. 4) chronometers onto an absolute timescale.

Angrites SAH99555 and NWA1296 contain highly radiogenic Pb (Table 1) that is more radiogenic than was previously determined for bulk meteorites. This renders these meteorites particularly suitable to high-precision Pb–Pb dating as uncertainties from laboratory blank, terrestrial contamination and choice of initial Pb become less important with increasingly radiogenic Pb. Data for whole-rock fragments of SAH99555, NWA1296 and the intensively acid-washed pyroxene from SAH99555 yield an isochron age of 4.5662 ± 0.0001 Gyr, with no assumption about initial Pb (Fig. 1). This is the oldest and most precise absolute age obtained for igneous rocks from our Solar System—just 1.0 ± 0.6 Myr younger than the currently accepted minimum age of the Solar System (4.5672 ± 0.0006 Gyr) obtained by Pb–Pb dating of calcium–aluminium-rich inclusions (CAIs) from undifferentiated (that is, chondritic) meteorites⁷. In relative terms, it represents the most precise absolute age determined on any material.

Mg isotopes were measured to high precision on the angrites to search for anomalies in ²⁶Mg abundances resulting from the decay of now-extinct ²⁶Al ($t_{1/2} = 730,000$ yr), which decays to ²⁶Mg. Angrite whole-rock fragments have ²⁶Mg excesses of 0.032‰ to 0.012‰ that are resolvable from matrix-matched basalt standards from the Earth, the Moon and Mars, as well as bulk carbonaceous and ordinary chondrites (Table 2). These data show, for the first time, that bulk samples of basaltic rocks from differentiated planetesimals have small ²⁶Mg excesses with respect to inner Solar System materials. Minerals from SAH99555 do not define an isochron and it is evident that the feldspar separate has partially re-equilibrated with low Al/Mg phases, most probably the fine-grained olivine with which it forms graphic intergrowths⁸. Combining the feldspar and bulk analysis for SAH99555 yields an age 5.6 ± 0.3 Myr after CAI formation, in excellent agreement with other ²⁶Al–²⁶Mg age data for (purer) feldspar separates from this angrite (5.7 ± 0.4 Myr; ref. 11). By reference to the absolute Pb–Pb age for CAIs⁷, the feldspar age (4.5616 ± 0.0007 Gyr) is much younger than the angrite Pb–Pb age and suggests that the feldspar records a late cooling or resetting age.

However, it is possible to calculate ²⁶Al–²⁶Mg model ages for angrite magmatism from bulk angrite analyses based on the assumption that ²⁶Al was homogeneously distributed in the young Solar System, and that the angrite parent body (APB) was not characterized by an unusual Al/Mg ratio or Mg isotopic composition compared to other Solar System materials. By reference to the initial ²⁶Al abundance of CAIs^{12,13}, high-precision Mg isotope data for bulk angrite fragments define model ages from 3.3 to 3.8 Myr after CAIs (Table 2), although uncertainties allow the four analysed angrites to have a range in ages as low as $\sim 100,000$ yr. The validity of these model ages is supported by three observations: (1) the olivine–kirschsteinite separate from SAH99555 with an Al/Mg ratio similar to bulk Solar System material has a ²⁶Mg abundance identical to terrestrial basalts; (2) the ²⁶Al–²⁶Mg model age for ADOR, a pyroxene-rich angrite with an unusually low Al/Mg ratio, is identical to that of the other angrites; (3) high-precision Mg isotope data for chondrites and basalts from the Earth, the Moon and Mars provide no evidence for inner Solar System heterogeneity of ²⁶Al or Mg isotopes (Table 2).

The most straightforward interpretation of the angrite Pb data are that the whole-rock isotopic variations represent small variations in U/Pb ratios due to modal heterogeneity of the fragments and/or dissolution of phases induced by acid washing, and that the crystallization age is 4.5662 ± 0.0001 Gyr. It is unlikely that the coherent and very old isochron age defined by different fragments of SAH99555 and NWA1296 and clinopyroxene separate from the former would result from the presence of now-extinct ²⁴⁷Cm (which decays to ²³⁵U), particularly as high-precision uranium

¹School of Earth Sciences, Victoria University of Wellington, PO Box 600, Wellington, New Zealand. ²Geological Institute, Øster Voldgade 10, DK-1350, Denmark. ³Geological Museum, Øster Voldgade 5–7, DK-1350, Denmark. ⁴Department of Geological Sciences, The University of Texas at Austin, 1 University Station C1100, Austin, Texas 78712-02, USA.

isotope studies have thus far failed to detect $^{238}\text{U}/^{235}\text{U}$ variations in bulk meteorites¹⁴. It is also improbable that our isochron dates an extreme U/Pb fractionation event of angrite pre-cursor material that significantly pre-dates angrite crystallization, given that the clinopyroxene analysis falls on the isochron.

Our angrite Pb–Pb isochron age is nearly 10 Myr older than pyroxene model Pb–Pb ages from the angrites ADOR, LEW86010 and D'Orbigny^{9,10} and suggests that angrites have variably experienced slow cooling or secondary thermal events, as is clear from the discrepant ^{53}Mn – ^{53}Cr and Pb–Pb mineral ages of D'Orbigny^{10,15} and Mg-isotope data for feldspar from angrites. ADOR and LEW86010 have relatively equilibrated textures and lack chemical zoning in some minerals, and D'Orbigny has a complex texture that implies multistage crystallization and metasomatic overprints^{8,16,17}. This contrasts with SAH99555, which has a relatively simple sub-ophitic igneous texture and strongly zoned mineral phases, and suggests that ADOR, LEW86010 and D'Orbigny underwent slower cooling, or more extensive thermal or metasomatic overprinting than SAH99555 in the APB. The apparently disturbed internal (that is, mineral) isotope systematics in nearly all igneous meteorites has important implications for the mapping of the ^{53}Mn – ^{53}Cr chronometer onto an absolute timescale, because this has used a combined Pb–Pb and ^{53}Mn – ^{53}Cr study of minerals from angrites⁵. However,

our study indicates the short-lived ^{53}Mn – ^{53}Cr and ^{182}Hf – ^{182}W chronometers can best be mapped onto an absolute timescale through whole-rock ^{53}Mn – ^{53}Cr and ^{182}Hf – ^{182}W angrite data or an internal isochron for SAH99555.

^{26}Mg excesses in bulk angrites reported here result in a calculated time difference between the formation of angrites and CAIs from carbonaceous chondrites of 3.3–3.8 Myr, in excellent agreement with the time difference between CAI and angrite formation that can be calculated from initial Sr isotope data^{11,18}. This time difference is greater than the difference in absolute Pb–Pb ages for the angrites and CAIs from the carbonaceous chondrite Efremovka⁷ (1.0 ± 0.6 Myr). Although this may reflect Allende CAIs being older than those from Efremovka, the Pb–Pb ages for Efremovka CAIs may simply not reflect the time of primary CAI formation as recorded by Al/Mg fractionation. The young CAI Pb–Pb ages could result from: (1) thermal overprints in the protoplanetary disk before incorporation in the CV chondrite parent body, which ^{26}Al – ^{26}Mg chondrule dating¹² constrains to have accreted at least 2 Myr after primary CAI formation, or (2) the common Pb component of the leached CAI fragments not being sufficiently precisely constrained by the data regressed to yield the ages in ref. 7. For example, if the CAI acid leachates most reliably represent the common Pb component in the CAIs (dominated by Pb introduced during terrestrial exposure

Table 1 | U–Pb isotopic results for standards and angrite samples

Sample	$^{204}\text{Pb}/^{206}\text{Pb}$ ($\pm 2\sigma$)	$^{207}\text{Pb}/^{206}\text{Pb}$ ($\pm 2\sigma$)	$^{208}\text{Pb}/^{206}\text{Pb}$ ($\pm 2\sigma$)	U (p.p.b.)	Pb (p.p.b.)	μ	Pb–Pb age (Myr)
Standard data							
SRM981 ($n = 9$)	0.059021 $\pm$ 13	0.914888 $\pm$ 37	2.16783 $\pm$ 13	–	–	–	–
Recommended values	0.059026	0.914901	2.16779	–	–	–	–
SRM983 ($n = 7$)	0.0003650 $\pm$ 18	0.071204 $\pm$ 06	0.013611 $\pm$ 19	–	–	–	–
SRM983 ($n = 9$)	0.0003633 $\pm$ 72	0.071201 $\pm$ 13	0.013601 $\pm$ 18	–	–	–	–
SRM983 ($n = 6$)	0.0003643 $\pm$ 37	0.071209 $\pm$ 36	–	–	–	–	–
NIST values	0.000371 $\pm$ 19	0.071201 $\pm$ 40	0.013619 $\pm$ 24	–	–	–	–
Th29 ($n = 6$)	0.055726 $\pm$ 22	0.859442 $\pm$ 44	2.09517 $\pm$ 21	–	~150	–	–
Recommended values	0.055729	0.859428	2.09519	–	–	–	–
JB-2 ($n = 7$)	0.054514 $\pm$ 40	0.848372 $\pm$ 60	2.08681 $\pm$ 21	–	~4,000	–	–
Recommended values	0.054515	0.848360	2.08676	–	–	–	–
Angrite data (cold 2 M HCl wash, 1 $\times$ 5 min)							
SAH99555 WR with surficial coating	0.0188200 $\pm$ 46	0.699521 $\pm$ 21	1.362674 $\pm$ 27	–	–	–	4,539.58 $\pm$ 0.08
Angrite data (cold 2 M HCl wash, 2 $\times$ 10 min)							
SAH99555 WR1	0.0007499 $\pm$ 84	0.627309 $\pm$ 24	1.162857 $\pm$ 40	29.25	76.93	1,223	4,564.77 $\pm$ 0.10
Leachate	0.0016790 $\pm$ 59	0.630805 $\pm$ 13	0.901430 $\pm$ 19	–	–	–	–
SAH99555 WR2	0.000946 $\pm$ 11	0.628139 $\pm$ 21	1.195657 $\pm$ 29	40.18	89.86	1,153	4,564.65 $\pm$ 0.12
Leachate	0.0018957 $\pm$ 54	0.631679 $\pm$ 17	0.908139 $\pm$ 21	–	–	–	–
SAH99555 WR3	0.000593 $\pm$ 12	0.626717 $\pm$ 28	1.190886 $\pm$ 43	24.94	71.62	1,431	4,565.04 $\pm$ 0.14
Leachate	0.0024765 $\pm$ 92	0.634252 $\pm$ 22	0.919434 $\pm$ 27	–	–	–	–
SAH99555 PX1	0.001113 $\pm$ 20	0.628859 $\pm$ 36	1.655911 $\pm$ 76	14.03	49.76	719.3	4,564.58 $\pm$ 0.22
Leachate	0.0418003 $\pm$ 33	0.792227 $\pm$ 17	0.848613 $\pm$ 14	–	–	–	–
Angrite data (cold 2 M HCl wash, 5 [WR] or 10 [pyroxene] $\times$ 10 min)							
SAH99555 WR4	0.0004134 $\pm$ 61	0.626106 $\pm$ 24	1.219479 $\pm$ 35	26.79	73.26	2,172	4,565.49 $\pm$ 0.08
(TIMS)	0.0004080 $\pm$ 40	0.62605 $\pm$ 31	–	–	–	–	4,565.41 $\pm$ 0.72
Leachate 1 & 2	0.0051557 $\pm$ 40	0.645565 $\pm$ 18	0.983099 $\pm$ 20	–	–	–	–
Leachate 3 & 4	0.000633 $\pm$ 16	0.627180 $\pm$ 35	0.885109 $\pm$ 40	–	–	–	–
Leachate 5	0.000503 $\pm$ 92	0.62629 $\pm$ 15	0.88354 $\pm$ 16	–	–	–	–
SAH99555 WR5	0.0003105 $\pm$ 65	0.625720 $\pm$ 21	1.339255 $\pm$ 28	20.53	58.72	2,883	4,565.66 $\pm$ 0.08
(TIMS)	0.0003168 $\pm$ 32	0.62579 $\pm$ 31	–	–	–	–	4,565.75 $\pm$ 0.71
SAH99555 PX2	0.000993 $\pm$ 31	0.628367 $\pm$ 57	2.141715 $\pm$ 118	8.759	30.02	957.6	4,564.69 $\pm$ 0.35
(TIMS)	0.001020 $\pm$ 10	0.62861 $\pm$ 31	–	–	–	–	4,564.98 $\pm$ 0.73
NWA1296 WR1	0.0007116 $\pm$ 22	0.627177 $\pm$ 13	1.078556 $\pm$ 15	26.36	96.57	895.4	4,564.87 $\pm$ 0.04
D'Orbigny WR1	0.0208978 $\pm$ 39	0.701790 $\pm$ 25	1.459173 $\pm$ 29	35.43	195.1	23.88	4,518.08 $\pm$ 0.09

Pb isotope methods modified from ref. 25 are described in the Supplementary Information. Except where indicated (TIMS), all data were acquired by multiple-collector inductively coupled plasma mass spectrometry (MC-ICPMS) using sample-standard bracketing methods to correct for instrumental mass bias. However, one set of SRM983 data ($n = 6$) and small aliquots of SAH99555 WR4, WR5 and PX2 were also analysed for Pb isotopes by ion counter peak jumping using thermal ionization mass spectrometry (TIMS). WR, whole rock. Model ages calculated from Pb isotope data determined on aliquots of these three samples by TIMS are identical to those calculated from MC-ICPMS data. Analyses of unradiogenic Pb from leachates of SAH99555 fragments and a piece of SAH99555 with a yellow coating developed during its exposure in the desert define an array distinct from the isochron shown in Fig. 1, $^{238}\text{U}/^{204}\text{Pb}$.

and/or extensive sample preparation), the Efremovka CAI for which the most precise Pb isotope data are available (E60) defines an age of 4.5695 ± 0.0004 Gyr (mean square weighted deviation, $\text{MSWD} = 1.7$) when data are normalized to the standard values obtained here. This corresponds to a time interval between angrite and CAI formation of 3.3 ± 0.4 Myr and in agreement with our results. Integration of our Pb–Pb and ^{26}Al – ^{26}Mg angrite chronometry with ^{26}Al – ^{26}Mg systematics for CAIs^{12,13} suggests that CAIs formed at $\geq 4.5695 \pm 0.0002$ Gyr. This revised, older, minimum age of the Solar System is in accordance with less precise constraints from other short-lived chronometers^{19–21}.

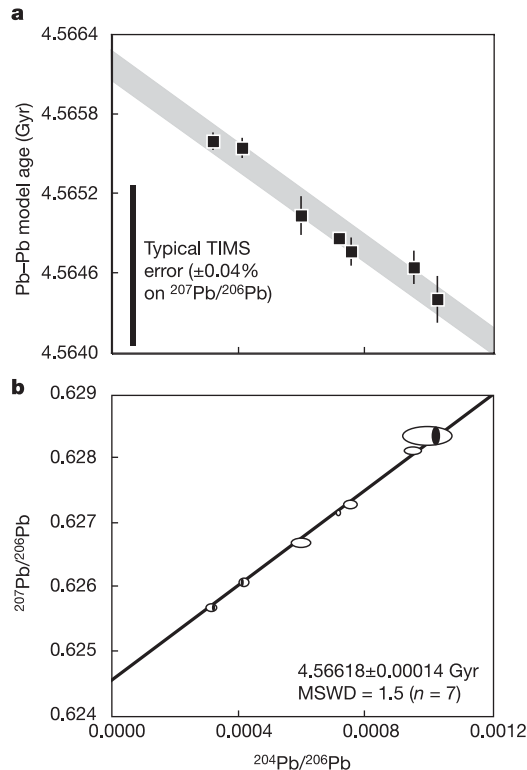


Figure 1 | Pb–Pb age and isotope data for angrites. **a**, Pb–Pb model ages for angrites versus $^{204}\text{Pb}/^{206}\text{Pb}$. Model ages increase with the proportion of radiogenic Pb (that is, decreasing $^{204}\text{Pb}/^{206}\text{Pb}$), as primordial Pb (PAT, ref. 27) does not appear to represent the initial angrite Pb. Therefore model Pb–Pb ages should not be used, even from meteorites with very radiogenic Pb, to decipher fine-scale information on early Solar System chronology. For example, the extreme U/Pb and early formation of the APB means that the quoted Pb–Pb model ages for pyroxene from the angrites ADOR and LEW86010 (ref. 9) can no longer be considered accurate as compared to the precision (± 0.4 Myr) of the dates—the actual pyroxene cooling or resetting ages must be slightly (ADOR; 1–2 Myr) or significantly (LEW86010; > 10 Myr) younger than the quoted PAT-model ages of 4.5578 ± 0.0004 Gyr. Model ages in this figure are calculated from MC-ICPMS data and, for three samples, from combining the more precise $^{204}\text{Pb}/^{206}\text{Pb}$ ratios obtained by TIMS with the MC-ICPMS $^{207}\text{Pb}/^{206}\text{Pb}$ data. Error bars are 2σ . The grey band has a vertical width equal to about $\pm 100,000$ yr. Also shown for reference is the uncertainty on model ages that these samples would have if the analyses had been performed by TIMS: that is, a $^{207}\text{Pb}/^{206}\text{Pb}$ error of $\pm 0.04\%$ corresponding to a ± 0.6 Myr age uncertainty. **b**, Pb–Pb isochron diagram for angrites. Fragments of SAH99555, NWA1296 and the intensively acid-washed clinopyroxene from SAH99555 define isochron ages of 4.56618 ± 0.00019 Gyr ($\text{MSWD} = 1.3$; MC-ICPMS data) and 4.56618 ± 0.00014 Gyr ($\text{MSWD} = 1.5$; using the more precise TIMS $^{204}\text{Pb}/^{206}\text{Pb}$ determinations for three samples). A whole-rock analysis of ADOR subjected to acid washing also lies on the isochron (not shown, ref. 27). Open white error ellipses, MC-ICPMS data; filled black error ellipses, combined MC-ICPMS $^{207}\text{Pb}/^{206}\text{Pb}$ and TIMS $^{204}\text{Pb}/^{206}\text{Pb}$ data. Error ellipses are 2σ .

The APB accreted and melted while ^{26}Al was present and sufficiently extant to induce melting and simple thermal modelling constrains accretion of the APB to ≥ 4.5680 Gyr (Fig. 2). The angrite crystallization and, in particular, APB accretion ages are older than nearly all precise absolute and relative ages for chondrules from undifferentiated or chondritic meteorites^{7,12,22,23} (Fig. 2). In this comparison, it is critical to appreciate that the youngest chondrule ages from undifferentiated or chondrite parent bodies provide a maximum age for their accretion. These constraints for timing of chondrite parent body accretion are analogous to those for a sedimentary rock from Earth where the youngest detrital mineral or rock fragments must define the maximum depositional age of the sediment, not the oldest components. Conversely, angrite magmatism provides a minimum age for APB accretion and formation of differentiated planetesimals. Thus, planetesimals like the APB must have accreted before chondrite parent bodies—despite the fact that chondrites are conventionally viewed as being the most primitive meteorites from first generation planetesimals. In this respect, our conclusions are in excellent agreement with high-precision W-isotope data for iron meteorites that indicate chondrite parent bodies accreted after those represented by some metallic cores of differentiated planetesimals²⁴. However, as some chondrules formed as early as CAIs¹², differentiated planetesimals may have ultimately accreted from material comparable to chondrites (that is, chon-

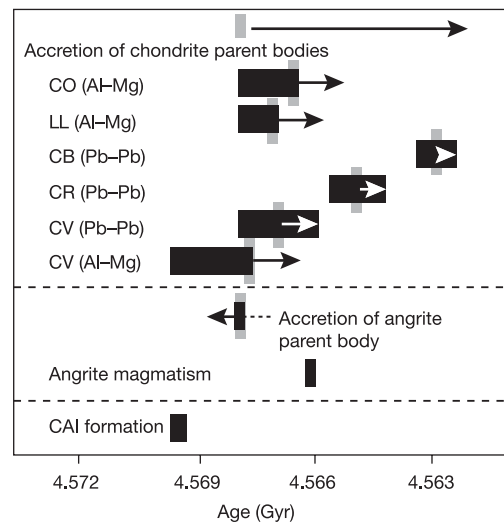


Figure 2 | Timing of angrite magmatism and parent body (APB) accretion compared with upper age limits for accretion of chondrite parent bodies. The youngest chondrule ages from different classes (CO, LL, CB, CR, CV) of undifferentiated chondrites constrain the maximum age for accretion of their parent bodies, because the youngest (undisturbed) ages for constituents of undifferentiated planetesimals must predate parent-body accretion. The timing of APB accretion calculated from thermal modelling overlaps or, in nearly all cases, predates the precise Pb–Pb ages and youngest Al–Mg ages for chondrules (refs 7, 12, 22, 23) from each chondrite parent body suggesting differentiated planetesimals accreted before chondrite parent bodies. The CAI age shown on the figure is calculated from the combined Al–Mg and Pb–Pb chronometry of angrites presented here and is older than the absolute age of ref. 7 (see text for discussion). The APB accretion age was calculated from a thermal model assuming negligible heat loss, instantaneous accretion, a relatively low peak temperature of 1,500 K and heating by only ^{26}Al . Changing the model parameters to more realistically take into account heat loss, a finite accretion period and higher peak temperatures would yield even older accretion ages for the APB. The black-filled rectangles encompass either the error on Pb–Pb ages or the range of ^{26}Al – ^{26}Mg ages for chondrules from individual chondrites. Arrows denoting the timing of chondrite parent body accretion extend from the mean age (Pb–Pb ages) or, in the case of ^{26}Al – ^{26}Mg ages, from the youngest chondrule ages that must predate accretion.

Table 2 | Al-Mg isotopic results for standards and angrite samples

Sample	²⁷ Al/ ²⁴ Mg	δ ²⁶ Mg (‰)	δ ²⁵ Mg (‰)	± 2σ	δ ²⁶ Mg* (‰)	± 2σ	n	ΔT (CAIs) (Myr)
Inner Solar System data								
BHVO-1 (Earth)	-	-0.0853	-0.0441	0.0206	-0.0024	0.0050	8	-
BCR-2 (Earth)	-	-0.1850	-0.0946	0.0181	-0.0007	0.0045	15	-
BIR-1 (Earth)	-	-0.2908	-0.1478	0.0144	0.0008	0.0059	16	-
NWA032 (Moon)	-	-0.2052	-0.1057	0.0145	0.0014	0.0045	9	-
EETA79001 (Mars)	-	-0.0727	-0.0406	0.0053	0.0019	0.0078	8	-
Allende CV3 chondrite	-	-0.3642	-0.1849	0.0297	-0.0032	0.0056	8	-
Allende chondrules	~0.1	-0.3955	-0.2024	0.0434	-0.0020	0.0097	3(9)	-
Murchison CM2 chondrite	-	-0.2759	-0.1440	0.0202	0.0039	0.0058	8	-
Julesburg L3 chondrite	-	-0.0140	-0.0066	0.0110	-0.0007	0.0072	8	-
Alfianello L6 chondrite	-	0.0584	0.0316	0.0230	-0.0044	0.0050	8	-
Angrite data								
SAH99555 WR1a	1.89	0.1819	0.0788	0.0207	0.0344	0.0075	7	3.24 ^{+0.24} _{-0.20}
SAH99555 WR1b	1.89	0.0193	-0.0097	0.0212	0.0338	0.0133	8	3.26 ^{+0.49} _{-0.33}
SAH99555 WR2	1.86	-0.0797	-0.0578	0.0141	0.0298	0.0067	13	3.38 ^{+0.25} _{-0.20}
SAH99555 WRs	1.88	-	-	-	0.0321	0.0047	3	3.31 ^{+0.16} _{-0.14}
SAH99555 olivine	0.086	-0.3566	-0.1826	0.0227	0.0027	0.0057	9	-
SAH99555 pyroxene	1.31	-0.0728	-0.0433	0.0150	0.0091	0.0051	15	-
SAH99555 feldspar	16.2	0.0979	0.0160	0.0118	0.0630	0.0097	4	5.55 ^{+0.34} _{-0.26}
NWA1296 WR1	1.75	-0.1770	-0.0999	0.0162	0.0169	0.0063	11	3.90 ^{+0.46} _{-0.32}
NWA1296 WR2	1.86	-0.1772	-0.1008	0.0228	0.0232	0.0078	9	3.64 ^{+0.40} _{-0.29}
NWA1296 WRs	1.81	-	-	-	0.0194	0.0049	2	3.80 ^{+0.29} _{-0.23}
D'Orbigny WR	1.86	-0.0914	-0.0596	0.0127	0.0268	0.0066	11	3.49 ^{+0.28} _{-0.22}
Angra dos Reis WR	0.929	-0.0241	-0.0181	0.0158	0.0122	0.0047	10	3.52 ^{+0.44} _{-0.31}

Mg isotope methods are described in the Supplementary Information. Because the relatively low temperature of Mg diffusion in anorthite²⁶ may have facilitated redistribution of Mg among minerals after crystallization, both whole-rock angrite fragments and minerals from SAH99555 were analysed to high precision. The feldspar age was calculated by combining the feldspar and whole-rock analysis of SAH99555. Mg isotope data for angrite fragments were combined with a bulk Solar System ²⁷Al/²⁴Mg = 0.10 ± 0.02 and the terrestrial Mg isotope composition to calculate initial abundances of ²⁶Al and the time of formation with respect to CAIs^{12,13}. In these cases, ΔT (CAIs) are model ages for basaltic magmatism on the APB, reflecting the increase in Al/Mg ratio produced by formation of basaltic magmas from a parent body with chondritic ²⁷Al/²⁴Mg of 0.10. Mg isotope data for the bulk angrite fragments yields model initial ²⁶Al abundances (²⁶Al/²⁷Al)₀ = 2.5 to 1.6 × 10⁻⁶. The initial ²⁶Al/²⁷Al of CAIs (5.83 ± 0.11 × 10⁻⁵) was recalculated from the data of ref. 12 using modified ²⁷Al/²⁴Mg ratios (the values published in ref. 12 are systematically too high by a factor of 1.11) and also the data for bulk CAIs from ref. 13. By reference to this initial ²⁶Al/²⁷Al of CAIs, angrite magmatism took place 3.3–3.8 Myr after CAIs. Use of a supra-canonical initial ²⁶Al/²⁷Al for CAIs to calculate ²⁶Al-²⁶Mg ages is further supported by the *in situ* laser ablation CAI data of ref. 13, which yields a CAI initial ²⁶Al/²⁷Al ≈ 6 × 10⁻⁵ when the data are treated in the same way as in ref. 12. n, number of analyses.

drules), although planetesimal melting has erased evidence for this. Relatively late accretion of chondrite parent bodies after ²⁶Al and ⁶⁰Fe abundances had decayed to low levels > 2 Myr after CAI formation (that is, less than 4.567 Gyr ago) would have prevented sufficient heating to melt their parent bodies.

Received 21 December 2004; accepted 1 June 2005.

- Lissauer, J. J. Timescales for planetary accretion and the structure of the protoplanetary disc. *Icarus* **69**, 249–265 (1987).
- Wasserburg, G. J. Isotopic abundances: inferences on solar system and planetary evolution. *Earth Planet. Sci. Lett.* **86** (The 1986 Crafoord Lectures), 129–173 (1987).
- Yin, Q. *et al.* A short timescale for terrestrial planet formation from Hf–W chronometry of meteorites. *Nature* **418**, 949–952 (2002).
- Kleine, T., Mezger, K., Münker, C., Palme, H. & Bischoff, A. ¹⁸²Hf-¹⁸²W isotope systematics of chondrites, eucrites, and martian meteorites: Chronology of core formation and early mantle differentiation in Vesta and Mars. *Geochim. Cosmochim. Acta* **68**, 2935–2946 (2004).
- Lugmair, G. W. & Shukolyukov, A. Early solar system timescales according to ⁵³Mn-⁵³Cr systematics. *Geochim. Cosmochim. Acta* **62**, 2863–2886 (1998).
- Urey, H. C. The cosmic abundances of potassium, uranium, and thorium and the heat balances of the Earth, the Moon, and Mars. *Natl Acad. Sci. USA* **41**, 127–144 (1955).
- Amelin, Y., Krot, A. N., Hutcheon, I. D. & Ulyanov, A. A. Lead isotopic ages of chondrules and calcium-aluminium-rich inclusions. *Science* **297**, 1678–1683 (2002).
- Mittlefehldt, D. W., Killgore, M. & Lee, M. T. Petrology and geochemistry of D'Orbigny and SAH99555, and the origin of angrites. *Meteorit. Planet. Sci.* **37**, 345–369 (2002).
- Lugmair, G. W. & Galer, S. J. G. Age and isotopic relationships among the angrites Lewis Cliff 86010 and Angra dos Reis. *Geochim. Cosmochim. Acta* **56**, 1673–1694 (1992).
- Jagoutz, E. *et al.* Cm?–U–Th–Pb isotopic evolution of the D'Orbigny angrite. *66th Annu. Meteorit. Soc. Meeting abstr.* 5148 (2003).

- Nyquist, L. E., Shih, C. Y., Wiesmann, H. & Mikouchi, T. Fossil ²⁶Al and ⁵³Mn in D'Orbigny and Sahara 99555 and the timescale for angrite magmatism. *Lunar Planet. Sci. Conf. XXXIV*, abstr. 1388 (2003).
- Bizzarro, M., Baker, J. A. & Haack, H. Mg isotope evidence for contemporaneous formation of chondrules and refractory inclusions. *Nature* **431**, 275–278 (2004).
- Young, E. D. *et al.* Supra-canonical ²⁶Al/²⁷Al and the residence time of CAIs in the solar protoplanetary disk. *Science* **308**, 223–227 (2005).
- Stirling, C. H., Halliday, A. N. & Porcelli, D. In search of live ²⁴⁷Cm in the early solar system. *Geochim. Cosmochim. Acta* **69**, 1059–1071 (2005).
- Glavin, D. P., Kubny, A., Jagoutz, E. & Lugmair, G. W. Mn–Cr isotope systematics of the D'Orbigny angrite. *Meteorit. Planet. Sci.* **39**, 693–700 (2004).
- Kurat, G. *et al.* D'Orbigny: A non-igneous angritic achondrite? *Geochim. Cosmochim. Acta* **68**, 1901–1921 (2004).
- Mittlefehldt, D. W. in *Treatise of Geochemistry* (ed. Davis, A. M.) Vol. 1, Ch. 11, 291–324 (Elsevier-Pergamon, Oxford, 2004).
- Halliday, A. N. & Porcelli, D. In search of lost planets—the paleocosmochemistry of the inner solar system. *Earth Planet. Sci. Lett.* **192**, 545–559 (2001).
- Gilmour, J. D. & Saxton, J. M. A time-scale of formation of the first solids. *Phil. Trans. R. Soc. Lond. A* **359**, 2037–2048 (2001).
- Lugmair, G. W. & Shukolyukov, A. Early solar system events and timescales. *Meteorit. Planet. Sci.* **36**, 1017–1026 (2001).
- Nyquist, L. *et al.* Manganese-chromium formation intervals for chondrules from the Bishunpur and Chainpur meteorites. *Meteorit. Planet. Sci.* **36**, 911–938 (2001).
- Amelin, Y., Krot, A. & Twelker, E. Pb isotopic age of the CB chondrite Gujba, and the duration of the chondrule formation interval. *Geochim. Cosmochim. Acta* **68**, abstr. E958 (2004).
- Kita, N. T. *et al.* Constraints on the origin of chondrules and CAIs from short-lived and long-lived radionuclides. *Workshop on Chondrites and Protoplanetary Disks abstr.* 9064 (2004).
- Kleine, T., Klaus, M., Herbert, P., Scherer, E. & Muenker, C. A new chronology for asteroid formation in the early solar system based on ¹⁸²W systematics. *Eos* **85**(47), Fall Meeting Suppl. abstr. P31C–04 (2004).

25. Baker, J., Peate, D., Waight, T. & Meyzen, C. Pb isotopic analysis of standards and samples using a ^{207}Pb - ^{204}Pb double spike and thallium to correct for mass bias with a double-focusing MC-ICPMS. *Chem. Geol.* **211**, 275–303 (2004).
26. LaTourrette, T. & Wasserburg, G. J. Mg diffusion in anorthite: implications for the formation of early solar system planetesimals. *Earth Planet. Sci. Lett.* **158**, 91–108 (1998).
27. Tatsumoto, M., Knight, R. J. & Allègre, C. J. Time differences in the formation of meteorites as determined from the ratio of lead-207 to lead-206. *Science* **180**, 1278–1283 (1973).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Financial support for this project was provided by the Danish Lithosphere Centre (funded by the Danish National Science Foundation). L. Labenne is thanked for his efforts in finding SAH99555 and providing our angrite samples. C. Stirling provided us with a pre-print of her U isotope study of meteorites. NASA supplied the Martian meteorite EETA79001. V. Fernandes provided us with lunar meteorite NWA032. Y. Amelin is thanked for his feedback on an earlier version of this paper.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.B. (joel.baker@vuw.ac.nz).

LETTERS

Meteoritic dust from the atmospheric disintegration of a large meteoroid

Andrew R. Klekociuk¹, Peter G. Brown², Dee W. Pack³, Douglas O. ReVelle⁴, W. N. Edwards², Richard E. Spalding⁵, Edward Tagliaferri³, Bernard B. Yoo³ & Joseph Zagari¹

Much of the mass of most meteoroids entering the Earth's atmosphere is consumed in the process of ablation. Larger meteoroids (>10 cm), which in some cases reach the ground as meteorites, typically have survival fractions near 1–25 per cent of their initial mass¹. The fate of the remaining ablated material is unclear, but theory suggests that much of it should recondense through coagulation as nanometre-sized particles². No direct measurements of such meteoric 'smoke' have hitherto been made³. Here we report the disintegration of one of the largest meteoroids to have entered the Earth's atmosphere during the past decade, and show that the dominant contribution to the mass of the residual atmospheric aerosol was in the form of micrometre-sized particles. This result is contrary to the usual view that most of the material in large meteoroids is efficiently converted to particles of much smaller size through ablation⁴. Assuming that our observations are of a typical event, we suggest that large meteoroids provide the dominant source of micrometre-sized meteoritic dust at the Earth's surface over long timescales.

United States Department of Defense space-based infrared sensors detected a large meteor in flight at an altitude of approximately 75 km at 12:07:20.975 UT (universal time) on 3 September 2004 at latitude 67.72° S, longitude 16.89° E. Space-based Department of Energy (DoE) visible light sensors also detected the fireball (Fig. 1). Subsequently, the emissive debris trail from the fireball was measured, extending from 56–18 km altitude and remained detectable at infrared wavelengths owing to solar scattering for over an hour. Two distinct disintegration features were visible along the path, at 32 km and 25 km altitude. We hereafter refer to this fragmentation region as 'ground zero'. Application of entry modelling^{5,6} of the light curve and trajectory data yielded initial mass estimates between $(0.6\text{--}1.9) \times 10^6$ kg, depending on the choice of various parameters in these models. Extrapolating a previously determined relation between optical energy and total energy⁷ derived for smaller bolides yields an estimate for total initial energy of $(5.4 \pm 0.4) \times 10^{13}$ J (equivalent to 13 ± 1 kilotons of exploding TNT) corresponding to a mass of $(0.65 \pm 0.05) \times 10^6$ kg. The original solar orbit of the body (Supplementary Table 1) is similar to near-Earth asteroids of the Aten group.

In addition to the satellite energy estimates, five infrasound stations detected acoustic gravity waves from the fireball, with the furthest detection being 13,000 km from ground zero. Using a recent calibration between wind-corrected observed fireball acoustic amplitudes and satellite yields⁸, a mean source energy of $(1.2 \pm 0.3) \times 10^{14}$ J (28 ± 6 kilotons of TNT) was obtained from signals detected at the four closest stations, equivalent to a mass of $(1.4 \pm 0.3) \times 10^6$ kg. The range of initial mass estimates from modelling and the optical and acoustic data correspond to a body

of diameter 7–10 m, assuming a mass density typical of chondritic meteorites⁹ of $\rho = 3,500 \text{ kg m}^{-3}$.

Some 7.5 h after the satellite observation, an anomalous 'cloud' was detected in the upper stratosphere by a polarization Rayleigh light detection and ranging (lidar) instrument at Davis station (68.6° S, 78.0° E) in Antarctica¹⁰ (Fig. 2). The cloud, which was directly related to the fireball event (as discussed below), was situated above the maximum height at which polar stratospheric clouds have previously been detected in September at Davis (~ 20 km altitude) and other similar-latitude sites^{11,12}. As determined by the lidar, local radiosondes and the Atmospheric Infrared Sounder (AIRS) satellite instrument¹³, temperatures in the vicinity of the cloud were near 240 K. This was ~ 55 K warmer than the expected frost-point for nitric acid trihydrate, which has the highest equilibrium temperature of solid polar stratospheric cloud constituents.

Using the location and time of the fireball, we examined three air parcel trajectory models to investigate aerosol dispersal in the context of the lidar observations. According to the best-fit model (from the Goddard Spaceflight Center¹⁴), air parcels from 32 km altitude at ground zero passed directly over Davis at 28.5 km altitude near 19:50 UT. This closely agrees with the onset of the strongest lidar backscatter shown in Fig. 2. Air from below 30 km altitude at ground

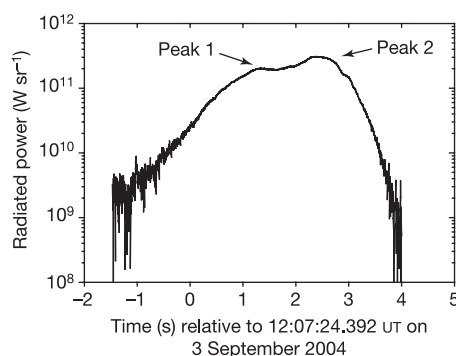


Figure 1 | Optical light curve for the 03 Sep 2004 fireball as measured by DoE space-based sensors. The signal was approximately 5.5 s in duration and exhibited two peaks due to discrete fragmentation events. These are assumed to be correlated with the features along the infrared track at altitude 32 km (located at 67.68° S, 18.00° E) and at 25 km altitude (67.67° S, 18.17° E). Intersection of the projected path with the Earth was at 67.64° S, 18.83° E (WGS-84 ellipsoid). The maximum radiated power of the event was $3.08 \times 10^{11} \text{ W sr}^{-1}$, and the total radiated energy was $7.26 \times 10^{12} \text{ J}$ (assuming that the fireball radiated as a 6,000-K blackbody). At its brightest, the bolide had an absolute visual magnitude of $M_v = -24$.

¹Space and Atmospheric Sciences, Australian Antarctic Division, Kingston, Tasmania 7050, Australia. ²Department of Physics and Astronomy, University of Western Ontario, London, Ontario N6A 3K7, Canada. ³The Aerospace Corporation, 2350 E. El Segundo Blvd, El Segundo, California 90245-4691, USA. ⁴Los Alamos National Laboratory, PO Box 1663, MS J577, Los Alamos, New Mexico 87545, USA. ⁵Sandia National Laboratory, Org. 5740, MS 0973, PO Box 5800, Albuquerque, New Mexico 87185, USA.

zero was modelled as passing more than 100 km north of Davis, and this suggests that the debris observed by lidar is from the upper infrared fragmentation event, but not the lower event. Based on the observed durations of the cloud features and using inferred wind fields, we estimate that the lower limit on the zonal extent of the trail at a given altitude was ~ 75 km as it passed Davis.

We modelled the arrival time–altitude profile of sedimenting particles at Davis (Fig. 3) to examine the potential range of particle sizes¹⁵. The main front of the lidar cloud features is well represented by the arrival of particles with radii of ~ 5 μm or less. In addition to the features shown in Fig. 2, narrow (and weak) aerosol layers with a mean backscatter ratio of ~ 0.05 above their immediate background, and that were tilted in the arrival time–altitude plane, were observed by lidar at lower altitudes. As indicated in Fig. 3, the upper layer had behaviour consistent with the sedimentation of particles with radii of ~ 10 – 20 μm originating near 30 km altitude at ground zero. The lower layer appears to represent the leading edge of an ensemble of different plumes of large particles.

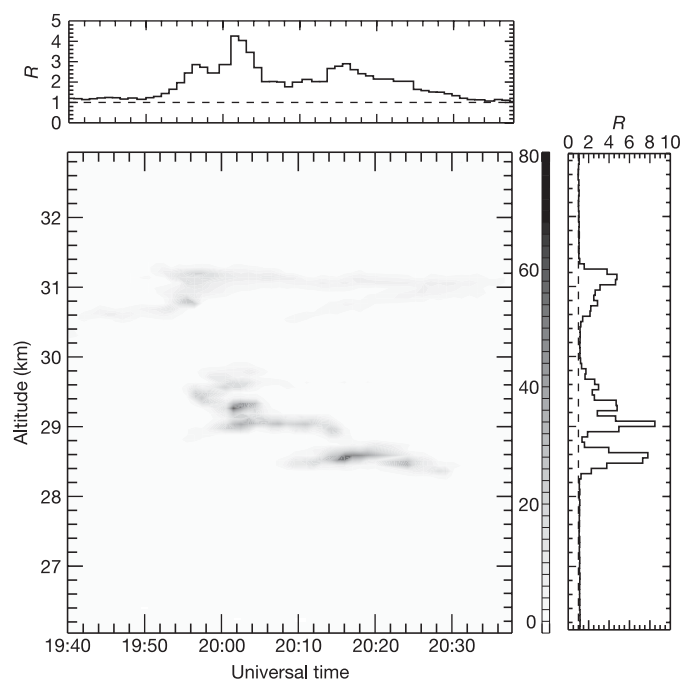


Figure 2 | The first unambiguous detection of lidar backscatter from the dust trail of a large meteoroid. Plotted is the lidar-derived total backscatter ratio R at a wavelength of 532 nm for a segment of the Davis observations on 3 September 2004, showing discrete structures detected between 19:38 UT and 20:37 UT. Here $R = 1 + \beta_m/\beta_a$, where β_m and β_a are the sums of the molecular and aerosol backscatter coefficients for orthogonally polarized signal components, respectively. The lidar retrieval was calibrated using the molecular number density inferred from AIRS satellite data¹³. Also shown are mean vertical and horizontal cross-sections for R . To investigate the time–altitude profile of aerosols detected by the lidar, three models of atmospheric transport were considered: the Goddard Spaceflight Center (GSFC) model¹⁴, the Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLOT) model from the National Oceanic and Atmospheric Administration, and the British Atmospheric Data Centre Trajectory Service. Of these models, only HYSPLIT evaluated particle sedimentation. All of the models clearly showed that for altitudes near 32 km, ground zero was directly upwind of Davis, with a strong (up to ~ 95 m s^{-1}) dominantly zonal flow linking the two sites. The speed of the flow accounted for the time difference between the bolide and lidar detections to better than 7%. In addition, the models suggested that the main cloud features represented suspended or weakly sedimenting particles (that is, descending by less than a few kilometres per day). The best agreement with the observations was obtained with the GSFC model using GSFC Data Assimilation Office (DAO) meteorological fields and diabatic corrections.

The lidar depolarization ratio¹⁶ as a function of backscatter ratio (Fig. 4) shows that the clouds were dominated by non-spherical (solid) aerosols. The lidar data also show that these aerosols were optically thin. We could only place an upper limit of 0.01 km^{-1} on the mean aerosol optical depth over the altitude range 28–31.5 km. This property was used together with the observed frequency distributions of depolarization ratio and backscatter ratio (in particular, the clustering into two groups shown in Fig. 4) to constrain the parameter space of effective particle radius, number density, shape and composition. This was done through comparison with scattering calculations^{17,18} for randomly oriented irregular spheroids comprising six cosmic dust analogue materials^{19–21} for particle radii up to 5 μm ; meteoric aerosol, glassy olivine, glassy pyroxene, silicon, iron and iron oxide. We found that the simplest match to the observed range and distribution of depolarization ratios was exhibited by olivine and pyroxene (materials found in chondritic meteorites²⁰) for effective radii of between ~ 0.3 and ~ 1.1 μm and nearly spherical shapes (shape factors between 0.9 and 1.1).

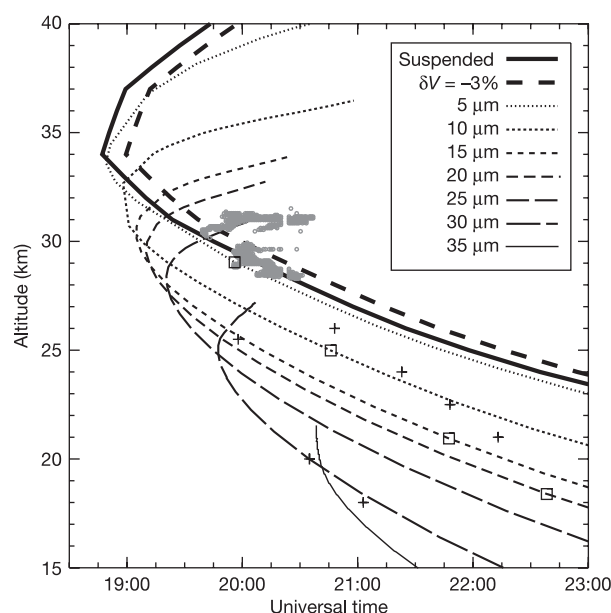


Figure 3 | Comparison between the time–altitude characteristics of the lidar-detected aerosols and the predictions of a sedimentation model. The model evaluated trajectories for spherical particles along the lagrangian path between ground zero and Davis using interpolated United Kingdom Meteorological Office (UKMO) Stratospheric Assimilated Data¹⁵. The plotted lines show the arrival time–altitude loci obtained from the model for the longitude of Davis, assuming a particle mass density of $3,500 \text{ kg m}^{-3}$. Overlaid are lidar scattering features (grey) having $R > 2$ (see Fig. 2). The crosses refer to weak aerosol features described in the text. Open squares show where particles of different radii released from ground zero at 30 km altitude would lie. The bold dashed line shows the case of suspended particles for a wind speed 3% lower than its mean value, to indicate one side of the expected spread in arrival times due to the influence of atmospheric gravity waves. The UKMO wind speed reached a peak value of $\sim 95 \text{ m s}^{-1}$ near 34 km altitude. The vertical gradient in the wind speed below the peak was $4 \text{ m s}^{-1} \text{ km}^{-1}$ and this accounted for the majority of the arrival time–altitude dispersion of the leading cloud fronts evident in Fig. 2. The cloud features above ~ 30 km appear more complex and extended than those at lower altitudes, and this may be related to shearing in the wind profile near 34 km, or produced by the fireball itself. An additional possibility is that the upper layers represent relatively heavy particles from higher altitudes (35 km to at least 60 km, the top of the UKMO model) that had fallen to this region. However, this association is problematic because the size distribution of the particles would need to be extremely narrow because of the small vertical extent of this cloud layer as a function of time.

Olivine gave the best agreement between the limiting depolarization values δ_1 for the two aerosol associations (Groups 1 and 2, shown in Fig. 4) and peaks in the calculated depolarization at effective radii r of 0.41 ± 0.12 and $0.98 \pm 0.29 \mu\text{m}$, respectively. Using lidar-derived mean backscatter coefficients and theoretical backscatter cross-sections we estimated effective particle-number densities N for the two groups as $(6.0 \pm 2.7) \times 10^6 \text{ m}^{-3}$ (Group 1) and $(1.7 \pm 0.4) \times 10^6 \text{ m}^{-3}$ (Group 2).

Taking the values of N , together with r and mass density ρ , and the inferred dimensions of the lidar-detected cloud (zonal $\sim 75 \text{ km}$, meridional $\sim 200 \text{ km}$ from trajectory modelling, vertical $\sim 3 \text{ km}$ from Fig. 2), we estimate the total mass of the aerosols as $(1.1 \pm 0.3) \times 10^6 \text{ kg}$. This value compares favourably with the meteoroid mass from the satellite and airwave data given above. We note that this mass estimate probably relates only to part of the debris cloud (that is, the upper fragmentation event) and hence is a lower limit. Further trajectory and scattering modelling is required to refine these values. We also note that that iron and iron oxide gave reasonable depolarization values at radii less than $\sim 0.3 \mu\text{m}$ for highly

non-spherical particle shapes. However, the inferred optical depth for these aerosols was significantly larger than that deduced from the lidar observations.

Our measurements suggest that a substantial fraction of the total ablated mass from large (metre-sized) chondritic bodies entering the atmosphere is deposited as $\sim$ micrometre-sized dust. This dust is likely to have atmospheric residence times of weeks to months. Micrometre-sized aerosols play a crucial role in climate forcing, through direct (radiative) and indirect (cloud nucleation) effects, as well as in ozone depletion through heterogeneous reactions. Aerosols with radii between about $0.05 \mu\text{m}$ and $1 \mu\text{m}$ scatter the most light per unit mass, and tend to have the longest atmospheric residence times²². The conventional view is that the background meteoritic aerosol flux is of low significance in climate forcing primarily because nanometre-sized particles dominate the size spectrum²³. In light of our findings, this view requires further investigation.

If most of the mass from large ($>0.1 \text{ m}$) meteoroid disintegrations is reduced to micrometre-sized particles, this process would dominate the influx to the Earth's surface of extraterrestrial dust at these sizes. Indeed, the mass flux from primary meteoroids of $1 \mu\text{m}$ and smaller is a few hundred tonnes per year, which is less than the mass delivered by sizes $>0.1 \text{ m}$ on century timescales²⁴. Although our measurements apply strictly to larger meteoroids disintegrating lower in the atmosphere (constituting a mass influx of $\sim 1 \text{ kt}$ per year), a substantial fraction of the total ablated mass of smaller meteoroids may also be partitioned into larger (micrometre versus nanometre) meteoric dust. Micrometre-sized ablation products cannot easily be measured individually in large numbers at the Earth's surface and so the mass distribution of meteoric material at such small sizes is largely unknown²⁵. In particular, models of atmospheric heating of interplanetary dust generally treat meteoroids as single ablating bodies, whereas much evidence suggests²⁶ that small (10^{-8} kg) meteoroids near the peak of the mass influx curve ablate as a collection of grains (dustballs), emphasizing the importance of fragmentation. In such a situation, the survival of a large fraction of the total incident meteoroid mass as micrometre-sized particles to the Earth's surface would be possible.

Received 16 March; accepted 1 June 2005.

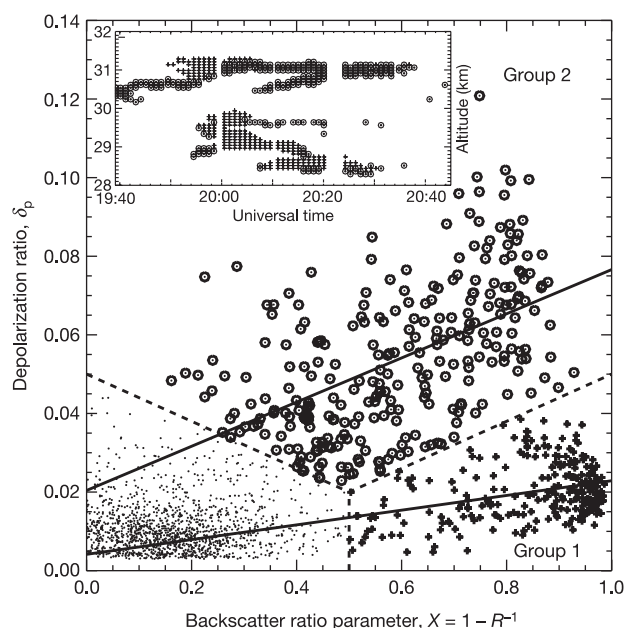


Figure 4 | Evidence that the lidar backscatter was due to non-spherical particles. Plotted is the lidar-derived depolarization ratio $\delta_p = S_p/S_s$ as a function of total backscatter parameter $X = 1 - R^{-1}$. Here S_p and S_s are the perpendicular and parallel polarized signal components, respectively, and R is the backscatter ratio. Divisions (dashed lines) have been assigned on the basis of the apparent clustering of points into distinct groups with the weighted linear regression fits to the rightmost (Group 1, crosses) and upper (Group 2; dotted circles) associations also shown (solid lines). The inset shows where points from Groups 1 and 2 lie in relation to the cloud. Clustered in the lower left-hand quadrant of the main plot are points representing the background aerosols that do not show a preferred association with the cloud features. Photon-counting noise, as well as gravity waves unresolved by the AIRS data, tends to scatter the points in the X -axis. For the non-spherical (solid) aerosols, Group 1 are characterized by low depolarization and high backscatter ratios and are associated with the leading edges of the strongest cloud layers. They have a limiting depolarization ratio δ_1 ($\lim_{X \rightarrow 1}$ of $2.3 \pm 0.4\%$). Group 2 particles have a wide range of depolarization and backscatter ratios, with $\delta_1 = 7.6 \pm 0.7\%$. The backscatter coefficient for Group 1 was mean $\beta_a = (2.351 \pm 0.003) \times 10^{-7} \text{ m}^{-1} \text{ sr}^{-1}$ and total $\beta_a = (5.831 \pm 0.008) \times 10^{-5} \text{ m}^{-1} \text{ sr}^{-1}$, and for Group 2 mean $\beta_a = (4.31 \pm 0.01) \times 10^{-8} \text{ m}^{-1} \text{ sr}^{-1}$ and total $\beta_a = (1.197 \pm 0.004) \times 10^{-5} \text{ m}^{-1} \text{ sr}^{-1}$. The Group 1 aerosols produced similar depolarization to those associated with an anomalous aerosol layer observed by lidar in the Arctic winter of 2000–2001 (ref. 27). This similarity strengthens the case for a meteoric origin for the Arctic layer.

1. Ceplecha, Z. et al. Meteor phenomena and bodies. *Space Sci. Rev.* **84**, 327–471 (1998).
2. Hunten, D. M., Turco, R. P. & Toon, O. B. Smoke and dust particles of meteoric origin in the mesosphere and stratosphere. *J. Atmos. Sci.* **37**, 1342–1357 (1980).
3. Jenniskens, P. Meteor induced chemistry, ablation products, and dust in the middle and upper atmosphere from optical spectroscopy of meteors. *Adv. Space Res.* **33**, 1444–1454 (2004).
4. Artemieva, N. A. & Shuvalov, V. V. Motion of a fragmented meteoroid through the planetary atmosphere. *J. Geophys. Res.* **106**, 3297–3309 (2001).
5. Ceplecha, Z., Spurný, P., Borovicka, J. & Kecklikova, J. Atmospheric fragmentation of meteoroids. *Astron. Astrophys.* **279**, 615–626 (1993).
6. ReVelle, D. O. in *Proc. Asteroids, Comets, Meteors (ACM 2002, International Conference, 29 July–2 August 2002, Berlin, Germany)* (ed. Warmbein, B.) ESA SP-500, 127–136 (ESA Publications, Noordwijk, Netherlands, 2002).
7. Brown, P., Spalding, R. E., ReVelle, D. O., Tagliaferri, E. & Worden, S. P. The flux of small near-Earth objects colliding with the Earth. *Nature* **420**, 294–296 (2002).
8. Edwards, W. N., Brown, P. & ReVelle, D. O. Bolide Energy Estimates from Infrasonic Measurements. *Earth, Moon and Planets* (Kluwer Academic, Dordrecht, in the press).
9. Britt, D. T. & Consolmagno, G. J. Stony meteorite porosities and densities: A review of the data through 2001. *Meteorit. Planet. Sci.* **38**, 1161–1180 (2003).
10. Klekociuk, A. R., Lambert, M. M., Vincent, R. A. & Dowdy, A. J. First year of Rayleigh lidar measurements of middle atmosphere temperatures above Davis, Antarctica. *Adv. Space Res.* **32**, 771–776 (2003).
11. Santacesaria, V., Mackenzie, A. R. & Stefanutti, L. A climatological study of polar stratospheric clouds (1989–1997) from LIDAR measurements over Dumont d'Urville (Antarctica). *Tellus B* **53**, 306–321 (2001).
12. Gobbi, G. P., Di Donfrancesco, G. & Adriani, A. Physical properties of stratospheric clouds during the Antarctic winter of 1995. *J. Geophys. Res.* **103**, 10859–10873 (1998).
13. Fetzer, E. et al. AIRS/AMSU/HSB Validation. *IEEE Trans. Geosci. Remote Sens.* **41**, 418–431 (2003).

14. Morris, G. A., Ziemke, J., Gleason, J. & Schoeberl, M. R. Trajectory mapping: A tool for satellite data validation. *J. Geophys. Res.* **105**, 17875–17894 (2000).
15. Swinbank, R. & O'Neill, A. A. Stratosphere-troposphere data assimilation system. *Mon. Weath. Rev.* **122**, 686–702 (1994).
16. Adachi, H., Shibata, T., Iwasaka, Y. & Fujiwara, M. Calibration method for the lidar-observed stratospheric depolarization ratio in the presence of liquid aerosol particles. *Appl. Opt.* **36**, 6578–6595 (2001).
17. Mishchenko, M. & Travis, L. D. Capabilities and limitations of current FORTRAN implementation of the T-Matrix method for randomly oriented, rotationally symmetric scatterers. *J. Quant. Spectrosc. Radiat. Transf.* **60**, 309–324 (1998).
18. Mishchenko, M. & Hovenier, J. W. Depolarization of light backscattered by randomly oriented nonspherical particles. *Opt. Lett.* **20**, 1356–1358 (1995).
19. Jager, C. et al. A database of optical constants of cosmic dust analogs. *J. Quant. Spectrosc. Radiat. Transf.* **79/80**, 765–774 (2003).
20. Rothman, L. S. et al. The HITRAN molecular database: Editions of 1991 and 1992. *J. Quant. Spectrosc. Radiat. Transf.* **48**, 469–507 (1992).
21. Lazzarin, M., di Martino, M., Barucci, M. A., Doressoundiram, A. & Florczak, M. Compositional properties of Near-Earth Asteroids: spectroscopic comparison with ordinary chondrite meteorites. *Astron. Astrophys.* **327**, 388–391 (1997).
22. Penner, J. E. et al. in *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change* (ed. Houghton, J. T. et al.) 289–348 (Cambridge Univ. Press, Cambridge, 2001).
23. Toon, O. B. & Farlow, N. H. Particles above the tropopause: Measurements and models of stratospheric aerosols, meteoric debris, nacreous clouds, and noctilucent clouds. *Annu. Rev. Earth Planet. Sci.* **9**, 19–58 (1981).
24. Lal, D. & Jull, A. J. T. Atmospheric cosmic dust fluxes in the range 10^{-4} to 10 centimeters. *Astrophys. J.* **576**, 1090–1097 (2002).
25. Karner, D. B. et al. Extraterrestrial accretion from the GISP2 ice core. *Geochim. Cosmochim. Acta* **67**, 751–763 (2003).
26. Fisher, A. A., Hawkes, R. L., Murray, I., Campbell, M. D. & LeBlanc, A. G. Are meteoroids really dustballs? *Planet. Space. Sci.* **48**, 911–920 (2000).
27. Gerding, M. et al. Observation of an unusual mid-stratospheric aerosol layer in the Arctic: possible sources and implications for polar vortex dynamics. *Ann. Geophys.* **21**, 1057–1069 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The Davis lidar observations are funded by the Department of the Environment and Heritage of the Australian Government. We thank D. McCormack and R. L. Hawkes for discussions. This work was supported in part by the Canada Research Chair programme, the Natural Sciences and Engineering Research Council of Canada and The Aerospace Corporation's Independent Research and Development programme.

Author Contributions A.R.K. developed the Davis lidar instrument, identified and analysed the bolide event in the lidar observations and co-wrote this Letter. P.G.B. helped interpret global infrasound network data, performed entry modelling and co-wrote this Letter. W.N.E. analysed global infrasound network data. D.O.R. performed entry modelling and helped interpret the global infrasound network. D.W.P., R.E.S., E.T. and B.B.Y. analysed Department of Defense and Department of Energy satellite data. J.Z. undertook initial and follow-up lidar observations, and assisted with data analysis.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.R.K. (andrew.klekociuk@aad.gov.au).

LETTERS

Ferroelectricity from iron valence ordering in the charge-frustrated system LuFe_2O_4

Naoshi Ikeda¹, Hiroyuki Ohsumi¹, Kenji Ohwada², Kenji Ishii², Toshiya Inami², Kazuhisa Kakurai³, Youichi Murakami⁴, Kenji Yoshii², Shigeo Mori⁵, Yoichi Horibe⁵ & Hijiri Kitô⁶

Ferroelectric materials are widely used in modern electric devices such as memory elements, filtering devices and high-performance insulators. Ferroelectric crystals have a spontaneous electric polarization arising from the coherent arrangement of electric dipoles¹ (specifically, a polar displacement of anions and cations). First-principles calculations^{2,3} and electron density analysis⁴ of ferroelectric materials have revealed that the covalent bond between the anions and cations, or the orbital hybridization of electrons on both ions, plays a key role in establishing the dipolar arrangement. However, an alternative model—electronic ferroelectricity⁵—has been proposed in which the electric dipole depends on electron correlations, rather than the covalency. This would offer the attractive possibility of ferroelectric materials that could be controlled by the charge, spin and orbital degrees of freedom of the electron. Here we report experimental evidence for ferroelectricity arising from electron correlations in the triangular mixed valence oxide, LuFe_2O_4 . Using resonant X-ray scattering measurements, we determine the ordering of the Fe^{2+} and Fe^{3+} ions. They form a superstructure that supports an electric polarization consisting of distributed electrons of polar symmetry. The polar ordering arises from the repulsive property of electrons—electron correlations—acting on a frustrated geometry.

Mixed valence material LuFe_2O_4 is a member of the RFe_2O_4 family, where R are rare-earth elements from Dy to Lu and Y (ref. 6). The crystal structure consists of the alternate stacking of triangular lattices of rare-earth elements, iron and oxygen. An equal amount of Fe^{2+} and Fe^{3+} coexists at the same site in the triangular lattice. Compared with the average iron valence of $\text{Fe}^{2.5+}$, Fe^{2+} and Fe^{3+} are considered as having an excess and a deficiency of half an electron, respectively. The coulombic preference for pairing of ‘oppositely’ signed charges (Fe^{2+} and Fe^{3+}) is considered to cause the degeneracy in the lowest energy for the charge configuration in the triangular lattice, similarly to the triangular antiferromagnetic Ising spins. Thus, RFe_2O_4 is considered to be a charge-frustrated system of triangular lattices.

Taking into account the charge frustration, a possible model of the superstructure of Fe^{2+} and Fe^{3+} in RFe_2O_4 has been proposed⁷. The postulated charge superstructure model is depicted in Fig. 1 for an adjacent iron triangular layer (W-layer). The competing interactions between frustrated charges are settled by this charge arrangement, similar to the stable configuration of Ising spins in triangular lattice⁸. The supercell, which is called a $\sqrt{3} \times \sqrt{3}$ structure, is enlarged by three times in the a - b plane along the (1 1 0) direction. A corresponding superstructure was detected by neutron diffraction. The presence of Bragg spots indexed as $(n/3 \ n/3 \ m + 1/2)$, where n and m are integers, was found below 330 K. Above 330 K, the spots

smear out to a Bragg line, indexed as $(n/3 \ n/3 \ l)$, where l takes a continuous value, which indicates the transformation of the three-dimensional (3D) ordering into a two-dimensional one. Below 250 K, the spin correlation develops as a ferrimagnetic ordering⁹.

Interestingly, the postulated charge structure allows the presence of a local electrical polarization, since the centres of Fe^{2+} (excess electron) and Fe^{3+} (electron deficiency) do not coincide in the unit cell of the superstructure. This indicates the possibility of ferroelectricity originating from the electron density modulation without a dipole of a cation and anion pair. Therefore, RFe_2O_4 is expected to be a ferroelectric material, reflecting the correlated nature of the electrons. As we show below, a large dielectric dispersion has been found in this material, in which the electron fluctuation has a central role in the ferroelectric domain boundary motion. Also, polarization switching in accordance with the external electric field has been reported. Though these facts suggested the existence of ferroelectricity, they were an insufficient basis on which to predict ferroelectricity by the ordered electrons, because the superstructure of Fe^{2+} and Fe^{3+} was estimated by neutron diffraction which is sensitive to lattice distortion but not to charge.

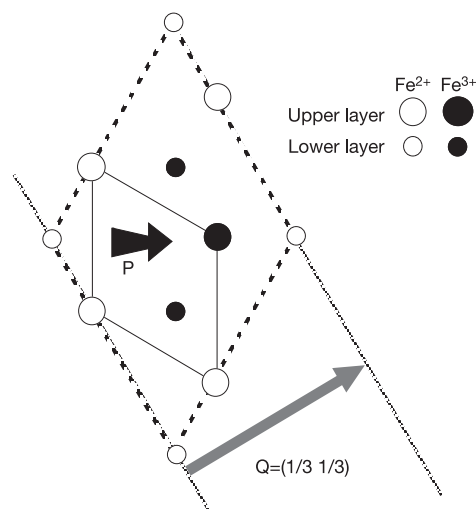


Figure 1 | The charge-ordering model of the double iron layer in RFe_2O_4 . The solid lines show the chemical unit cell. The dotted lines show the charge superlattice cell. The Fe ions in upper or lower layers are indicated with a large or a small circle, respectively. Fe^{2+} and Fe^{3+} are represented by open and filled circles, respectively. The polarization $\mathbf{P}$ is represented with a short, black arrow. A wave vector $\mathbf{Q}$ showing a charge wave is drawn with a long, grey arrow.

¹Japan Synchrotron Radiation Research Institute, SPring-8, Hyogo 679-5198, Japan. ²Synchrotron Radiation Research Center, SPring-8, Japan Atomic Energy Research Institute, Hyogo 679-5148, Japan. ³Japan Atomic Energy Research Institute, Ibaraki 319-1195, Japan. ⁴Department of Physics, Tohoku University, Sendai 980-8578, Japan. ⁵Department of Physical Science, Osaka Prefecture University, Sakai, Osaka 599-8531, Japan. ⁶National Institute of Advanced Industrial Science and Technology, Tsukuba 305-8561, Japan.

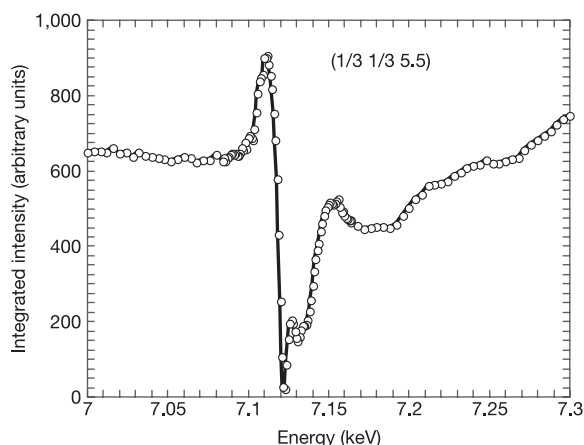


Figure 2 | X-ray energy dependence of the superlattice reflection (1/3, 1/3, 5.5) of a LuFe_2O_4 single crystal. Data have been corrected for the absorption effect. The peak and the minimum arise from the in-phase component of the anomalous atomic scattering factor for Fe^{2+} and Fe^{3+} , respectively

In order to clarify the existence of the superstructure of Fe^{2+} and Fe^{3+} , we performed a resonant X-ray scattering (RXS) experiment, which is the most sensitive technique for detecting the charge superstructure^{10–12}. The experiment on single crystal LuFe_2O_4 was made with a four-circle diffractometer at beamlines BL02B1 and BL22XU in SPring-8. A crystal sample grown by the floating zone melting method was cut to orient the c axis normal to the sample surface. The intensity of a superlattice spot ($n/3\ n/3\ m + 0.5$), where n and m are integers, was recorded as a function of the X-ray energy near the K-absorption edge of iron. The sample was cooled in a refrigerator down to 18 K, which is far below the transition temperature.

A typical result of the RXS experiment is displayed in Fig. 2. The absorption effect was corrected in the data. In the spectrum, a maximum and a minimum at around 7.113 keV and 7.120 keV and a background independent of X-ray energy were found. The extrema at 7.113 keV and 7.120 keV arise from the in-phase component of the anomalous atomic scattering factors of Fe^{2+} and Fe^{3+} , respectively. The corresponding K-absorption edges for both ions were confirmed by X-ray absorption near-edge structure measurements for LuCoFeO_4 and LuFeGaO_4 , which are isostructural to RFe_2O_4 but contain only Fe^{3+} and Fe^{2+} , respectively.

This result clearly indicates that the structure factor of this superlattice point is contributed by the positive atomic scattering factor of Fe^{3+} and the negative factor of Fe^{2+} . Therefore, we can conclude that the structure factor at this Bragg point arises from the 'difference' of atomic scattering factors for Fe^{2+} and Fe^{3+} . This is evidence for the formation of the long-range ordering of Fe^{2+} and Fe^{3+} with the $\sqrt{3} \times \sqrt{3}$ structure. This superlattice reflection appears below 330 K, which indicates the development of three-dimensional charge ordering or the occurrence of a Verwey transition. At the same time, spontaneous electric polarization appears below 330 K. This correspondence shows that the charge ordering is an order parameter of the electric polarization.

We demonstrated polarization switching in accordance with the external electric field detected by pyro-electric current observations of single-crystal LuFe_2O_4 (ref. 13). In the experiment, the sample was cooled down to 77 K under an electric field of $\pm 10\text{ kV cm}^{-1}$ along the c axis, and then the current flow from the sample was recorded upon heating without the electric field. The results showed that the direction of current flow depended on the sign of the cooling electric field below 350 K. We estimated the spontaneous polarization by the integral of the current from the sample. Figure 3 shows the obtained temperature dependence of the electric polarization. A large decrease

occurs as the temperature is increased around 250 K, which is the magnetic transition temperature, and at 330 K at which the superstructure of Fe^{2+} and Fe^{3+} appears. This shows that LuFe_2O_4 is a polar substance that can be switched by an external electric field, and that the ferroelectricity is developed by the polar arrangement of Fe^{2+} and Fe^{3+} .

The shoulder of the electric polarization at the magnetic transition temperature indicates the coupling of magnetization with electric polarization. In general, magnetic ordering hardly affects electric polarization¹⁴. But the mechanism described in this Letter, polarization formed by the polar arrangement of Fe^{2+} and Fe^{3+} , may allow such coupling because the coherent arrangement of spins on iron ions give rise to the magnetization. At present we have not investigated this phenomenon completely, but the shoulder suggests that the coherence length of the charge-ordered region or the polarization domain is connected to the development of exchange coupling of iron spins. Though the details are not yet clear, this shoulder demonstrates a potential multiferroic property of ferroelectricity caused by ordered electrons.

Concerning the ferroelectric properties of RFe_2O_4 , a characteristic large dispersion had been observed for low frequency alternating current (a.c.) dielectric constants. These are shown in Fig. 4a for LuFe_2O_4 . The response is the relaxation process expressed by Debye-type dispersion with the amplitude reaching 5,000. Large dielectric dispersion with an amplitude of the order of 10^4 is also observed in the iso-structural family ErFe_2O_4 (ref. 15). The dielectric dispersion has a common feature with the order–disorder type of ferroelectric materials, where motion of the ferroelectric domain boundary gives rise to the dispersion. These similarities indicate the presence of the ferroelectric domain and its boundary motion, giving rise to the dispersion.

An analysis of the low-frequency dispersion suggests the origin of this ferroelectricity. At a given temperature, a characteristic response frequency was found at a peak in the frequency variation of the imaginary part of the dielectric constant, ϵ'' . The temperature variation of the characteristic frequency obeys an Arrhenius relation:

$$f = f_0 \exp(-Q/kT) \quad (1)$$

where f is the characteristic frequency, kT is the thermal energy, f_0 is a prefactor and Q is the activation energy. Observations of the LuFe_2O_4 crystal along the c -axis are summarized in Fig. 4b, where $f_0 = 2.8 \times 10^{11}\text{ Hz}$ and $Q = 0.29\text{ eV}$. For comparison, the valence fluctuation frequencies of iron ions obtained by Mössbauer

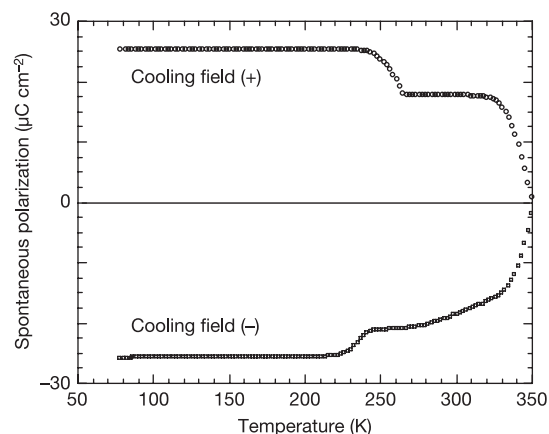


Figure 3 | Temperature variation of the electric polarization of LuFe_2O_4 . The plot is the integration of a pyro-electric current measurement. The current flow from the sample was recorded on heating after electric field cooling along the c axis. The direction of the electric polarization depends on the direction of electric field, which indicates that LuFe_2O_4 possesses macroscopic electric polarization.

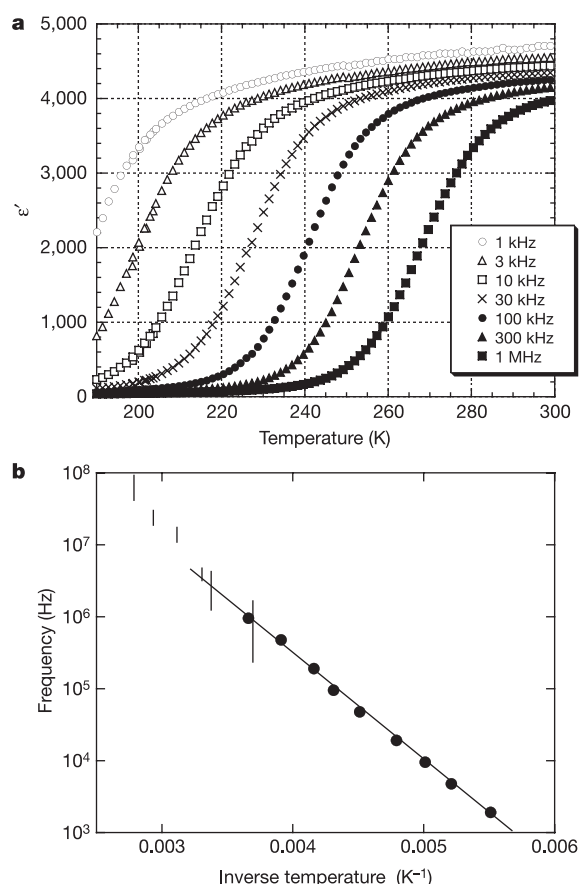


Figure 4 | The a.c. dielectric dispersion in an LuFe_2O_4 polycrystalline sample. **a**, The in phase component of the a.c. dielectric constants measured in the frequency range from 1 kHz to 1 MHz. For the observation at 1 kHz, the dielectric constant is more than 4,000 above 220 K. **b**, The temperature variation of the characteristic frequency of the dielectric dispersion. The Arrhenius relation is indicated by a full line. Thick vertical lines represents the iron valence fluctuating frequencies measured with Mössbauer spectroscopy¹⁶. Extrapolation shows that the electron hopping on the iron ion is the origin of the dispersion.

spectroscopy¹⁶ are marked by thick vertical lines in this figure. The extrapolation of the relation passes through the valence fluctuation frequency, indicating that the electron fluctuation of the iron ions plays a central role in the dielectric response and the motion of the ferroelectric domain boundary. This is consistent with the fact that the polarization arises from the electron ordering. The ferroelectric domain boundary movement proceeds with the exchange of electrons between Fe^{2+} and Fe^{3+} at the domain boundary.

These experimental findings—switching of electric polarization and large dielectric constant—are consistent with a criterion for the existence of ferroelectricity. The order parameter of the electric polarization is the ordering of Fe^{2+} and Fe^{3+} in an arrangement of polar symmetry. The dielectric dispersion shows a typical character of order–disorder type ferroelectric materials where the ferroelectric

domain boundary proceeds with electron exchange between Fe^{2+} and Fe^{3+} .

Thus we conclude that LuFe_2O_4 is a ferroelectric material, and that this property originates from the polar arrangement of electrons on Fe^{3+} . The electron arrangement arises from charge frustration on a triangular lattice. This arrangement of electrons is realized by the density modulation of d electrons, which is different from the premise of electronic ferroelectricity—it is assumed that in a certain kind of transition metal compound with a d - and f -electron system the balance between electron transfer and electron correlation gives rise to the polar arrangement of electrons. In the case of RFe_2O_4 , a particular condition, the charge frustration, leads to the same result: ferroelectricity. This ferroelectricity caused by the electron correlation offers great potential when designing future ferroelectric devices to be coupled or controlled with the degrees of freedom of electrons: charge, spin and orbital. Such properties may lead to a new multiferroic material. Also, the low activation energy of electron motion in this material, which suggests less coupling of polarization switching with the lattice distortion, may enable the development of a fatigue-free solid charge capacitor. These possibilities of new ferroelectric materials will be examined in future studies.

Received 2 May; accepted 11 July 2005.

1. Kittel, C. *Introduction to Solid State Physics* (Wiley, New York, 1995).
2. Cohen, R. E. Origin of ferroelectricity in perovskite oxides. *Nature* **358**, 136–138 (1992).
3. Săghi-Szabó, G., Cohen, R. E. & Krakauer, H. First-principles study of piezoelectricity in PbTiO_3 . *Phys. Rev. Lett.* **80**, 4321–4324 (1998).
4. Kuroiwa, Y. et al. Evidence for Pb–O covalency in tetragonal PbTiO_3 . *Phys. Rev. Lett.* **87**, 217601 (2001).
5. Portengen, T., Östreich, Th. & Sham, L. J. Theory of electronic ferroelectricity. *Phys. Rev. B* **54**, 17452–17463 (1996).
6. Kimizuka, N., Muromachi, E. & Siratori, K. in *Handbook on the Physics and Chemistry of Rare Earths* Vol. 13 (eds Gschneidner, K. A. Jr & Eyring, L.) 283–384 (Elsevier Science, Amsterdam, 1990).
7. Yamada, Y., Nohdo, S. & Ikeda, N. Incommensurate charge ordering in charge-frustrated LuFe_2O_4 system. *J. Phys. Soc. Jpn* **66**, 3733–3736 (1997).
8. Mekata, M. & Adachi, K. Magnetic structure of CsCoCl_3 . *J. Phys. Soc. Jpn* **44**, 806–812 (1978).
9. Funahashi, S. et al. Two-dimensional spin correlation in YFe_2O_4 . *J. Phys. Soc. Jpn* **53**, 2688–2696 (1984).
10. Materlik, G., Sparks, C. J. & Fischer, K. *Resonant Anomalous X-Ray Scattering* (North-Holland, Amsterdam, 1994).
11. Sasaki, S. Fe^{2+} and Fe^{3+} ions distinguishable by X-ray anomalous scattering: method and its application to magnetite. *Rev. Sci. Instrum.* **66**, 1573–1576 (1995).
12. Murakami, Y. et al. Direct observation of charge and orbital ordering in $\text{La}_{0.5}\text{Sr}_{1.5}\text{MnO}_4$. *Phys. Rev. Lett.* **80**, 1932–1935 (1998).
13. Ikeda, N. et al. Charge frustration and dielectric dispersion in LuFe_2O_4 . *J. Phys. Soc. Jpn* **69**, 1526–1532 (2000).
14. Hill, A. N. Why are there so few magnetic ferroelectrics? *J. Phys. Chem. B* **104**, 6694–6709 (2000).
15. Ikeda, N., Kohn, K., Kito, H., Akimitsu, J. & Siratori, K. Dielectric relaxation and hopping of electrons in ErFe_2O_4 . *J. Phys. Soc. Jpn* **63**, 4556–4564 (1994).
16. Tanaka, M., Siratori, K. & Kimizuka, N. Mössbauer study in RFe_2O_4 . *J. Phys. Soc. Jpn* **53**, 760–772 (1984).

Acknowledgements The authors express their gratitude to H. Suematsu and M. Takata for discussions.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to N.I. (ikedan@spring8.or.jp).

Catalytic enantioselective reactions driven by photoinduced electron transfer

Andreas Bauer¹, Felix Westkämper¹, Stefan Grimme² & Thorsten Bach¹

Photoinduced electron transfer is an essential step in the conversion of solar energy into chemical energy in photosystems I and II (ref. 1), and is also frequently used by chemists to build complex molecules from simple precursors². During this process, light absorption generates molecules in excited electronic states that are susceptible to accepting or donating electrons. But although the excited states are straightforward to generate, their short lifetimes makes it challenging to control electron transfer and subsequent product formation—particularly if enantiopure products are desired. Control strategies developed so far use hydrogen bonding, to embed photochemical substrates in chiral environments³ and to render photochemical reactions enantioselective through the use of rigid chiral complexing agents⁴. To go beyond such stoichiometric chiral information transmission, catalytic turnover is required⁵. Here we present a catalytic photoinduced electron transfer reaction that proceeds with considerable turnover and high enantioselectivity. By using an electron accepting chiral organocatalyst that enforces a chiral environment on the substrate through hydrogen bonding, we obtain the product in significant enantiomeric excess (up to 70%) and in yields reaching 64%. This performance suggests that photochemical routes to chiral compounds may find use in general asymmetric synthesis.

In a catalytic photochemical reaction, the catalyst acts as an antenna collecting the light and transferring it to the substrate via sensitization. Sensitization can occur by energy or electron transfer. The most successful enantioselective sensitizers rely on chirality transfer in a conformationally restricted exciplex⁶. The best results reported for a unimolecular reaction are 77% enantiomeric excess (77% e.e.; 20 mol% catalyst) on an analytical scale⁷ and for a bimolecular reaction 58% e.e. (15 mol% catalyst) at a product yield of 1% (ref. 8).

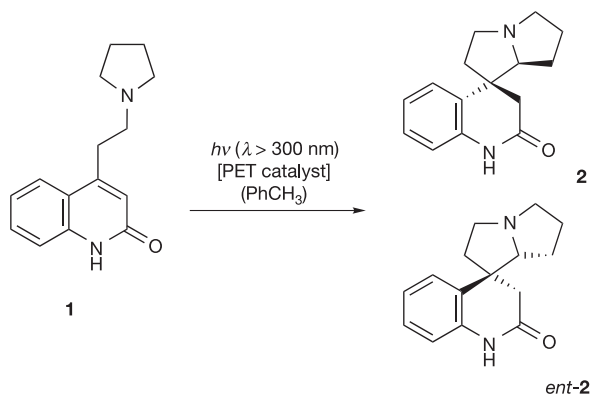


Figure 1 | PET-catalysed cyclization of the prochiral substrate **1** to the chiral pyrrolizidine **2** and its enantiomer *ent*-**2**.

The reaction we have devised for studying an enantioselective photoinduced electron transfer (PET) sensitization is depicted in Fig. 1. It is based on previously reported PET catalysed conjugate additions of α -amino alkyl radicals to enones that had been performed non-enantioselectively^{9,10}. It proceeds from substrate **1** (see Supplementary Scheme a and Supplementary Information page SI 4) to a chiral spirocyclic pyrrolizidine, which in an achiral environment is obtained as a mixture of **2** and its enantiomer, *ent*-**2**. The simple diastereoselectivity of the reaction is perfect. Only one diastereoisomer is formed, the configuration of which was proved by ¹H-NMR nuclear Overhauser effect (NOE) experiments.

In the presence of a catalyst, ultraviolet irradiation induces a PET from the amine to the photoexcited catalyst. Subsequent proton loss from the intermediate cation radical presumably leads to an α -aminoalkyl radical (see also Fig. 3), which adds intramolecularly to carbon atom C-4 of the quinolone. After the radical addition reaction, back electron transfer from the catalyst generates an enolate, which is eventually protonated to yield the products. As mentioned above, similar intramolecular⁹ and intermolecular¹⁰ addition reactions to enones are known. A suitable PET catalyst for these reactions was 4,4'-dimethoxybenzophenone (**3**, Fig. 2). By employing ketone **3** as a catalyst (10 mol%), the desired reaction **1** $\rightarrow$ **2**/*ent*-**2** proceeded in good yield (Table 1, entry 1) but of course without any enantioselectivity. The chiral catalyst **4** that we employed for enantioselective reactions has two key elements. First, it allows for binding of the substrate **1** by two hydrogen bonds at the bridgehead lactam. Second, it contains the catalytic benzophenone unit, which is bound to the 1,5,7-trimethyl-3-azabicyclo[3.3.1]nonan-2-one via a rigid oxazole. It therefore serves not only as a PET catalyst, but also as a stereocontrolling device inducing the desired enantiofacial differentiation in the cyclization step. Indeed, more flexible catalysts were far less successful in the attempted enantioselective reaction. The synthesis of compound **4** and its enantiomer *ent*-**4** was straightforward, based on our previous work (see Supplementary Scheme b and Supplementary Information pages SI 6–8). Experiments were conducted varying the catalyst loading and the substrate concentration. The results are summarized in Table 1.

Even with only 5 mol% catalyst, a reasonable product yield of 61% was obtained (Table 1, entry 2). This gives a calculated turnover number (12.2) that is unprecedented when considering that the product was formed with significant enantiomeric excess (20%). Upon raising the catalyst concentration, the reaction time decreased and the enantioselectivity increased (entries 3, 5, 6). The enantiomeric excess reached 70% for a catalyst loading of 30 mol%, which gave a calculated turnover number of 2.1 (entry 6). Note that these turnover numbers have not been corrected for any uncatalysed processes, even though a racemic background reaction evidently occurs and causes the decrease of e.e. upon decreasing the amount of catalyst (entries 6, 5, 3, 2). In fact, when quinolone **1** was irradiated in

¹Lehrstuhl für Organische Chemie I, Technische Universität München, Lichtenbergstr. 4, D-85747 Garching, Germany. ²Universität Münster, Organisch-Chemisches Institut, Corrensstr. 40, D-48149 Münster, Germany.

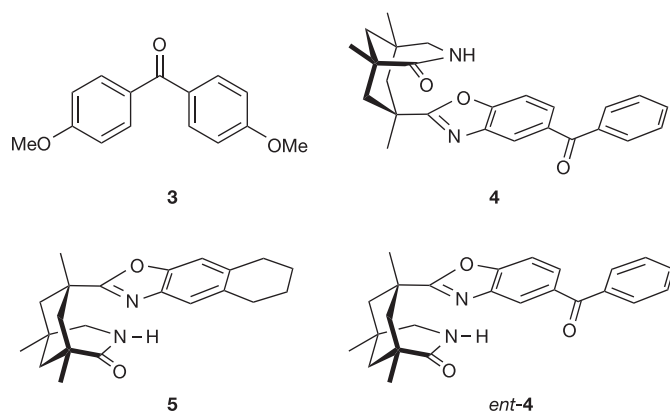


Figure 2 | Structures of the achiral PET-catalyst **3**, of the chiral enantiomeric PET-catalysts **4** and *ent*-**4**, and of the chiral complexing agent **5**.

the absence of a PET catalyst, the compound decomposed and 25% of racemic pyrrolizidine **2/ent-2** was isolated after 5 h. The uncatalysed reaction probably involves radical chain reactions as proposed in similar cyclizations¹⁰, and is increasingly suppressed in the presence of increasing amounts of photocatalyst **4**. (In the unlikely event that the background reactions occurred to the same degree in the presence of **4**, the corrected turnover number for entry 2 would be 7.2.)

If the other antipode of the catalyst was employed, the absolute configuration of the product was expectedly the opposite, that is, enantiomer *ent*-**2** prevailed (entry 4). The product configuration was tentatively assigned on the basis of known radical processes that occur in the presence of chiral complexing agent **5** (refs 11, 12). For comparison, a reaction was conducted in the presence of stoichiometric amounts of complexing agent **5** (entry 7), which has the same absolute configuration as catalyst *ent*-**4**. The product enantiomer *ent*-**2** was formed preferentially. Proof for the assignment was obtained by comparing the calculated specific rotation ($[\alpha]_D$) and circular dichroism (CD) spectra with experimental values^{13,14}. An enantiomerically pure compound with the given structure **2** was calculated to show a negative specific rotation with negative Cotton effects at 270, 235 and 210 nm and a positive Cotton effect at 255 nm (see Supplementary Information pages SI 10–13). Experimentally, enantiomer **2** obtained from the reactions with catalyst **4** was laevorotatory and exhibited the predicted Cotton effects in the CD spectra. The facial differentiation is explained in Fig. 3, assuming radical **6** as intermediate.

Mechanistic details of the PET catalysed reaction have not yet been elucidated. The single steps of the catalytic cycle will be studied in further experiments. The hydrogen atom transfer, for example, may in fact occur directly from the suggested ketyl radical and not via the proposed back electron transfer. The concept of chirality multiplication via a hydrogen-bonded catalyst like **4** should be successfully applicable to other photochemical reactions. Work along these lines is currently in progress in our laboratory.

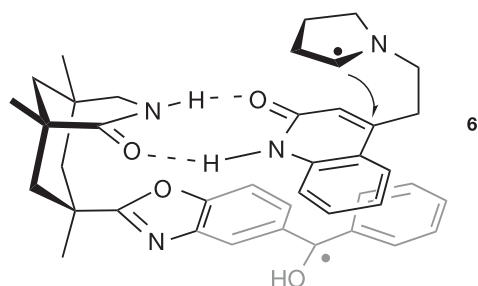


Figure 3 | Explanation for the facial differentiation in the PET-catalysed cyclization of the prochiral substrate **1** via radical intermediate **6**.

Table 1 | Enantioselective catalytic PET reactions of substrate **1** (see Figs 1 and 2)

Entry no	Catalyst	Equiv.*	Time (h)	Product	e.r.†	e.e.‡ (%)	Yield§ (%)
1	3	0.1	3.5	2/ent-2	50/50	—	71
2	4	0.05	5	2	60/40	20	61
3	4	0.1	2.5	2	69/31	38	55
4	<i>ent</i> - 4	0.1	3	<i>ent</i> - 2	31/69	38	52
5	4	0.2	2	2	77/23	54	57
6	4	0.3	1	2	85/15	70	64
7	3/5	0.1/1.2	2	<i>ent</i> - 2	14/86	72	39

* The reactions were carried out in deaerated toluene as the solvent at -60°C (irradiation source: Original Hanau TQ 150) and with a substrate concentration of 4 mM (see Supplementary Information page SI 5).

† The enantiomeric ratio (e.r.) was determined by ^1H -NMR shift experiments (see Supplementary Information page SI 9)¹⁵.

‡ The enantiomeric excess (e.e.) was calculated from the e.r. based on the uncertainty of the ^1H -NMR integration, the variance of e.e. data are estimated as $\pm 2\%$.

§ Yield of isolated product.

|| A stoichiometric amount (1.2 equiv.) of the chiral complexing agent **5** was added to the reaction mixture.

Received 5 April; accepted 27 June 2005.

- Nelson, N. & Ben-Shem, A. The complex architecture of oxygenic photosynthesis. *Nature Rev. Mol. Cell Biol.* **5**, 971–982 (2004).
- Schmoltdt, P., Rinderhagen, H. & Mattay, J. in *Molecular and Supramolecular Photochemistry* Vol. 9 (eds Ramamurthy, V. & Schanze, K. S.) 185–225 (M. Dekker, New York, 2003).
- Grosch, B. & Bach, T. in *CRC Handbook of Organic Photochemistry and Photobiology* (eds Horspool, W. & Lenci, F.) 61/1–61/14 (CRC Press, Boca Raton, 2004).
- Bach, T., Bergmann, H., Grosch, B. & Harms, K. Highly enantioselective intra- and intermolecular [2+2] photocycloaddition reactions of 2-quinolones mediated by a chiral lactam host: Host-guest interactions, product configuration, and the origin of the stereoselectivity in solution. *J. Am. Chem. Soc.* **124**, 7982–7990 (2002).
- Noyori, R. *Asymmetric Catalysis in Organic Synthesis* Ch. 1 (Wiley, New York, 1994).
- Inoue, Y. in *Molecular and Supramolecular Photochemistry* Vol. 11 (eds Inoue, Y. & Ramamurthy, V.) 129–177 (M. Dekker, New York, 2004).
- Hoffmann, R. & Inoue, Y. Trapped optically active (E)-cycloheptene generated by enantiodifferentiating Z-E photoisomerization of cycloheptene sensitized by chiral aromatic esters. *J. Am. Chem. Soc.* **121**, 10702–10710 (1999).
- Asaoka, S., Wada, T. & Inoue, Y. Microenvironmental polarity control of electron-transfer photochemistry. Enantiodifferentiating polar addition of 1,1-diphenyl-1-alkenes photosensitized by saccharide naphthalenecarboxylates. *J. Am. Chem. Soc.* **125**, 3008–3027 (2003).
- Jeon, Y. T., Lee, C.-P. & Mariano, P. S. Radical cyclization reactions of α -silyl amine α,β -unsaturated ketone and ester systems promoted by single electron transfer photosensitization. *J. Am. Chem. Soc.* **113**, 8847–8863 (1991).
- Bertrand, S., Hoffmann, N. & Pete, J.-P. Highly efficient and stereoselective radical addition of tertiary amines to electron-deficient alkenes—Application to the enantioselective synthesis of necine bases. *Eur. J. Org. Chem.*, 2227–2238 (2000).
- Bach, T., Aehtner, T. & Neumüller, B. Enantioselective Norrish-Yang cyclization reactions of N-(ω -oxoalkyl)substituted imidazolidinones in solution and in the solid state. *Chem. Eur. J.* **8**, 2464–2475 (2002).
- Aehtner, T., Dressel, T. & Bach, T. Hydrogen bond mediated enantioselectivity of radical reactions. *Angew. Chem. Int. Edn Engl.* **43**, 5849–5851 (2004).
- Polavarapu, P. L. Optical rotation: Recent advances in determining the absolute configuration. *Chirality* **14**, 768–781 (2002).
- Diedrich, C. & Grimme, S. Systematic investigation of modern quantum chemical methods to predict electronic circular dichroism spectra. *J. Phys. Chem. A* **107**, 2524–2539 (2003).
- Bergmann, H., Grosch, B., Sitterberg, S. & Bach, T. An enantiomerically pure 1,5,7-trimethyl-3-azabicyclo[3.3.1]nonan-2-one as ^1H -NMR shift reagent for the ee determination of chiral lactams, quinolones, and oxazolidinones. *J. Org. Chem.* **69**, 970–973 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the Deutsche Forschungsgemeinschaft and by the Fonds der Chemischen Industrie. We thank O. Ackermann for technical assistance and T. Straßner for discussions.

Author Contributions A.B. and F.W. contributed equally to the experimental part of the study. S.G. performed the DFT calculations.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.B. (thorsten.bach@ch.tum.de).

Base stacking controls excited-state dynamics in A-T DNA

Carlos E. Crespo-Hernández¹, Boiko Cohen¹ & Bern Kohler¹

Solar ultraviolet light creates excited electronic states in DNA that can decay to mutagenic photoproducts. This vulnerability is compensated for in all organisms by enzymatic repair of photodamaged DNA. As repair is energetically costly, DNA is intrinsically photostable. Single bases eliminate electronic energy non-radiatively on a subpicosecond timescale¹, but base stacking and base pairing mediate the decay of excess electronic energy in the double helix in poorly understood ways. In the past, considerable attention has been paid to excited base pairs². Recent reports have suggested that light-triggered motion of a proton in one of the hydrogen bonds of an isolated base pair initiates non-radiative decay to the electronic ground state^{3,4}. Here we show that vertical base stacking, and not base pairing, determines the fate of excited singlet electronic states in single- and double-stranded oligonucleotides composed of adenine (A) and thymine (T) bases. Intrastrand excimer states with lifetimes of 50–150 ps are formed in high yields whenever A is stacked with itself or with T. Excimers limit excitation energy to one strand at a time in the B-form double helix, enabling repair using the undamaged strand as a template.

Many DNA photolesions, including bipyrimidine photodimers, are formed from singlet excited states^{5–8}. Understanding DNA singlet excited-state dynamics is thus essential for determining how photolesions are formed and how their formation depends on base sequence and helix conformation. In double-stranded DNA, bases are organized horizontally in base pairs and vertically in base stacks. It is unknown which of these dual architectural motifs has the greatest influence on the nature and dynamics of singlet excited states. On the one hand, π -stacking of planar aromatic molecules facilitates the formation of excimer (excited dimer) states, in which excitation is shared between two closely spaced molecules. On the other hand, hydrogen bonds between paired bases suggest possible interstrand proton or hydrogen atom transfer². The experimental picture is ambiguous. Although excimers have been observed in diverse oligo- and polynucleotides^{9,10}, quantum yields of formation have never been measured. This makes it impossible to rule out alternative decay channels, including excited-state proton transfer. Gas-phase studies have detected such transfer in base pair mimics¹¹, and quenching of singlet states in polynucleotides at low temperature has been attributed to base pairing¹². However, the relevance of these results for room-temperature oligo- and polynucleotides is uncertain.

We recorded femtosecond transient absorption signals for defined-sequence oligonucleotides that differ in their extent of base stacking and base pairing in aqueous solution (Fig. 1). The DNA oligonucleotide (dA)₁₈ composed of 18 consecutive 2'-deoxyadenosine (dA) residues was investigated first. Bases in this model single strand are (mostly) stacked, but unpaired. Transient absorption at visible probe wavelengths by this 18-mer decays multiexponentially (Fig. 2) and resembles signals from the homopolymers poly(A) and

poly(dA)¹³. A rapid initial decay (time constant $\tau_1 = 0.80 \pm 0.16$ ps, all uncertainties are 2σ ; fitting parameters for all fits are in Supplementary Tables 1–4) is followed by much slower relaxation ($\tau_2 = 126 \pm 8$ ps), which is assigned to singlet excimers. Approximately 10% of the initial signal remains 1 ns after the pump pulse, indicating that these states decay on multiple timescales. In contrast, the mononucleotide adenosine-5'-monophosphate (AMP) shows only a subpicosecond decay ($\tau = 0.33 \pm 0.03$ ps). At a probe wavelength of 250 nm, where AMP and (dA)₁₈ absorb strongly, negative signals are observed. These bleaching signals monitor repopulation of the electronic ground state, S_0 . Importantly, the (dA)₁₈ transient at 250 nm decays in lock step with the signal at 570 nm at times greater than 10 ps (uppermost panel in Fig. 2). This demonstrates that the excimer state responsible for the long-time signal in (dA)₁₈ decays to S_0 , and not to a long-lived triplet state.

A fraction of excited states in (dA)₁₈ decay to S_0 by ultrafast internal conversion. Evidence for this monomer-like decay channel comes from absorption at ultraviolet (UV) probe wavelengths by vibrationally highly excited S_0 molecules, which cool by energy transfer to the solvent¹⁴. This results in the ~ 2 ps time constant seen in the (dA)₁₈ and AMP signals at 250 and 280 nm. The 280 nm signals in Fig. 2 exhibit consecutive kinetics consistent with the scheme in Fig. 1a: $S_{FC} \rightarrow S_0' \rightarrow S_0$, where S_{FC} and S_0' are the initial Franck–Condon excited state and vibrationally hot ground state, respectively. The greater magnitude of the short-time signal from AMP (Fig. 2, 280 nm panel) indicates that ultrafast internal conversion is substantially more important in the monomeric base than in (dA)₁₈.

The excimer formation yield in (dA)₁₈ can be estimated from the data in Fig. 2. From the fit amplitudes at 250 nm (Supplementary Table 1), 37% of the excited population is present as excimers 10 ps after the pump pulse. This number is a lower bound since states other than S_0 may absorb at 250 nm. In fact, the greater bleach for AMP compared to (dA)₁₈ (Fig. 2) is consistent with substantial excimer absorption at 250 nm. A preliminary analysis indicates that $\sim 65\%$ of the absorbed photons generate excimers in (dA)₁₈. This number is close to the fraction of stacked bases expected in (dA)₁₈ (79% of A bases are stacked in poly(dA); ref. 15), indicating that nearly every excitation of a stacked A-base decays to an excimer. This is the basis for the simple kinetic model of Fig. 1: every excitation in a base stack decays to an excimer, while every excitation of an unstacked base decays by internal conversion. Although more elaborate schemes can be envisioned for these heterogeneous, multichromophoric systems, this simple model adequately explains the present experimental results.

Unlike (dA)₁₈, excimers are not observed in (dT)₁₈, the single-stranded DNA 18-mer containing consecutive thymidine (dT) residues. Signals at 570 nm from (dT)₁₈ and its 5'-mononucleotide TMP agree within experimental uncertainty (Fig. 3), indicating that most excitations decay by ultrafast internal conversion. A slowly decaying

¹Department of Chemistry, The Ohio State University, 100 W. 18th Avenue, Columbus, Ohio 43210, USA.

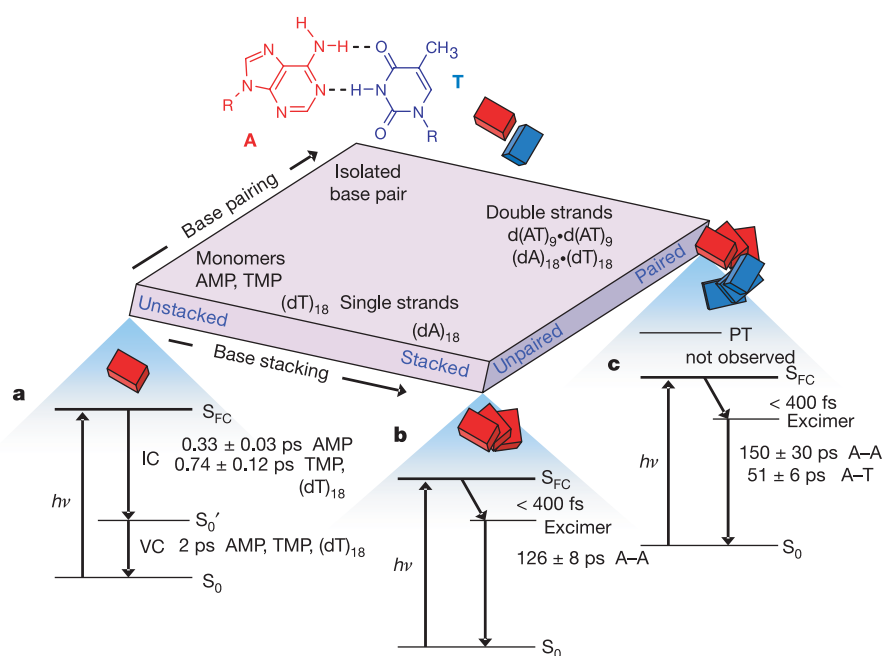


Figure 1 | DNA photophysical pathways as a function of base pairing and base stacking interactions. Model compounds containing adenine (A, red boxes) and thymine (T, blue boxes) are plotted at top in a two-dimensional space that qualitatively illustrates their degree of base-paired or base-stacked character. **a–c**, Approximate energies of excited singlet states in the DNA base monomers and oligonucleotides investigated. Lifetimes determined by femtosecond pump-probe measurements (see Figs 2–4) are shown beneath each state. UV light absorption transfers molecules from the electronic ground state, S_0 , to the optically allowed (that is, ‘bright’) Franck–Condon singlet state, S_{FC} . **a**, In the 5′-mononucleotides AMP and TMP, S_{FC} decays in < 1 ps by internal conversion (IC) to vibrationally excited ground states, S_0' , which re-equilibrate with the solvent by vibrational cooling (VC). **b**, Here, S_{FC} represents excitations localized on stacked bases in the single-stranded oligonucleotide, $(dA)_{18}$. These decay to excimer states on an ultrafast timescale, which is not resolved in these measurements. **c**, Even in the presence of base pairing, singlet excitations in A–A and A–T stacks are quenched to excimers with high efficiency, and there is no evidence for decay via a proton transfer state (PT).

bleach is observed in $(dT)_{18}$ but an excimer state can be ruled out since the same decay is seen in TMP. Instead, the slower decay ($\tau_2 = 103 \pm 18$ ps) at 253 nm is assigned to an unidentified singlet intermediate initially formed in 10–15% yield. Intersystem crossing to the triplet state ($\phi_{ISC} = 1.4 \times 10^{-2}$; ref. 16) accounts for most of the $\sim 3\%$ bleach seen in TMP at 500 ps. However, the triplet yield in $(dT)_{20}$ is an order of magnitude lower than in TMP (ref. 16), and we attribute the residual bleach in the oligonucleotide to the formation of thymine photodimers, which do not absorb significantly above 250 nm (ref. 5). Absorption spectra recorded before and after UV irradiation confirm that photodimers were formed in significant amounts in $(dT)_{18}$ (Supplementary Fig. 5). The bleach signal for $(dT)_{18}$ at 253 nm is larger than that for TMP at all delay times, possibly indicating that the thymine photodimer is formed on a subpicosecond timescale. Prompt formation would explain why photodimer formation is promoted by T–T base stacking^{17,18}.

Transients from $(dA)_{18} \cdot (dT)_{18}$ closely resemble those from $(dA)_{18}$, but have only 70–80% as much amplitude (Fig. 4a). Long-lived signal components in $(dA)_{18} \cdot (dT)_{18}$ are thus assigned to excimers in the A strand, which are not quenched by base pairing with T. Fewer A–A excimers are formed in the duplex because of absorption by the T strand. This decrease is offset partially by increased base stacking in the double-stranded oligonucleotide, which favours excimer formation. The absence of new absorption bands and residual bleaching at long times suggest that proton or hydrogen atom transfer is not a major deactivation channel. The lack of excited-state proton transfer in stacked A–T base pairs does not exclude this decay channel in isolated Watson–Crick base pairs, a subject addressed in recent publications^{3,4}. However, only ultrafast, monomer-like lifetimes were observed in a recent time-resolved study of A–T base pairs in the gas phase¹⁹. On the other hand, long-lived excited states have been detected in the stacked mixed dimer formed from 9-methyladenine and adenine in the gas phase²⁰.

Further evidence that base stacking governs singlet excited states in DNA base multimers is shown in Fig. 4b. The $(dA)_{18} \cdot (dT)_{18}$ and $d(AT)_9 \cdot d(AT)_9$ duplexes have identical numbers of A–T base pairs but differ in ‘vertical’ architecture. The long-lived signals from $d(AT)_9 \cdot d(AT)_9$ at 250 and 570 nm show that excimers are also formed

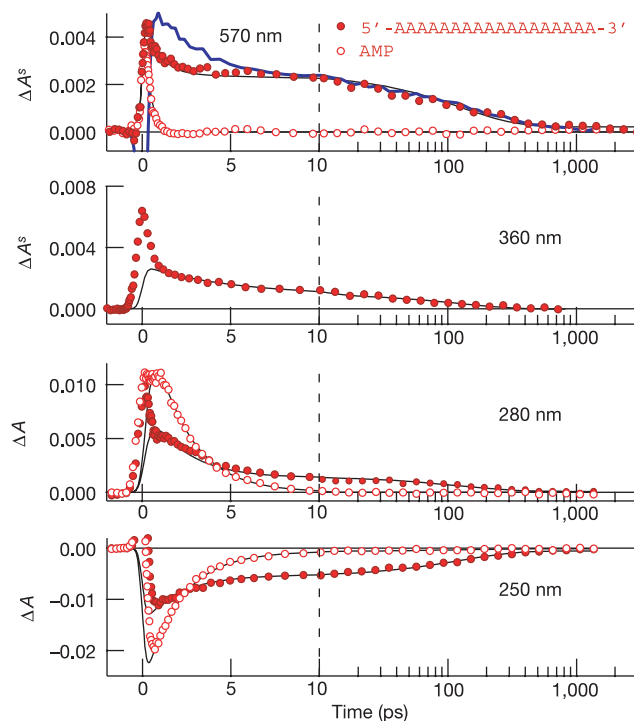


Figure 2 | Femtosecond pump-probe transients for the single-stranded oligonucleotide $(dA)_{18}$ (filled circles) and the 5′-mononucleotide AMP (open circles). Smooth curves are from a global fit described in Supplementary Table 1. Transient absorption (ΔA) at the indicated probe wavelength is plotted versus time delay between pump (266 nm) and probe pulses. At 570 nm, the AMP signal has been normalized to the maximum of the $(dA)_{18}$ signal. In other panels, signals were recorded in back-to-back experiments. The blue curve in the uppermost panel is the inverted $(dA)_{18}$ signal at 250 nm, scaled for comparison with the long-time signal at 570 nm. In this and later figures, the time axis is linear for delays of less than 10 ps, and logarithmic thereafter. Supplementary Figs 1–4 present the same data on linear time axes. Signals indicated by ΔA^s were corrected for two-photon ionization of the solvent.

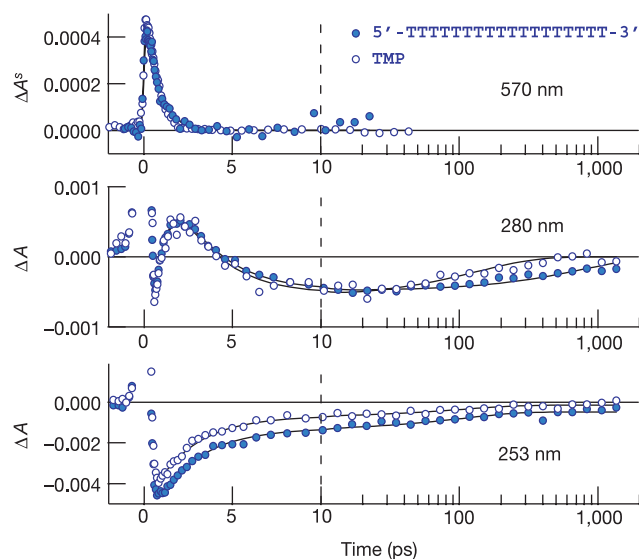


Figure 3 | Transient absorption versus time at several probe wavelengths for the single-stranded oligonucleotide $d(T)_{18}$ (filled circles) and its constituent 5'-mononucleotide, TMP (open circles). Excitation was at 266 nm. Smooth curves are from a global fit (see Supplementary Table 2).

in A–T base stacks in high yield, but are shorter-lived ($\tau = 51 \pm 6$ ps) than the A–A ones in $(dA)_{18}$ ($dT)_{18}$ ($\tau = 150 \pm 30$ ps). Still, base pairing effects are observable: the lower panel of Fig. 4b shows an increase in the A–T excimer lifetime in D_2O versus H_2O . This isotope effect, which is greatest in the slow time constant ($\tau_2 = 66 \pm 14$ ps in H_2O , 150 ± 30 ps in D_2O), demonstrates that hydrogen (deuterium) bonds in the double helix modulate the rate of excimer decay. We suggest that this modulation is greatest for excimers with significant charge transfer character. Coupling of proton and electron motions can significantly alter the driving force for charge recombination²¹. Charge resonance effects are important generally for aromatic excimers²², but are likely to be especially so in the A–T case owing to the different redox potentials of these bases. Consistent with this picture, no isotope effect was observed for A–A excimers in $(dA)_{18}$ ($dT)_{18}$.

The charge transfer character of DNA excimers also explains their low tendency to decay by photon emission. Thus, the radiative transition probability of the A–A excimer must be at least two orders of magnitude lower than the $^1\pi, \pi^*$ state of AMP to account for the modest fluorescence increase seen, for example, in adenine homopolymers compared to AMP (ref. 13). This 'dark' character explains why these long-lived states have gone undetected in femtosecond fluorescence up-conversion experiments²³. The low photolesion yields found in DNA despite high yields of long-lived excimers is probably due to the efficiency of excimer decay, which has significant charge recombination character. Interestingly, charge transport in duplex DNA is easier along strands, and interstrand hops are inhibited^{24,25}. These similarities suggest that a unified theory can describe both charge migration and the dynamics of singlet excited states in DNA.

In summary, oligo- and polynucleotides¹³ have dramatically different decay channels for singlet excitation energy compared to their constituent bases. Subpicosecond internal conversion—a hallmark of monomer excited states—is observed only in $(dT)_{18}$ and unstacked bases, whereas excitations in stacked A,T sequences decay overwhelmingly to intrastrand excimers. These results have important implications for DNA photochemistry. Although relatively long-lived excimers are formed in high yields in $(dA)_{18}$, $(dA)_{18}$ ($dT)_{18}$ and $d(AT)_9$ ($d(AT)_9$), each is significantly more photostable than $(dT)_{18}$ (Supplementary Fig. 5). Singlet state lifetimes are not simply correlated with the likelihood of photoreaction. Nonetheless, the strong

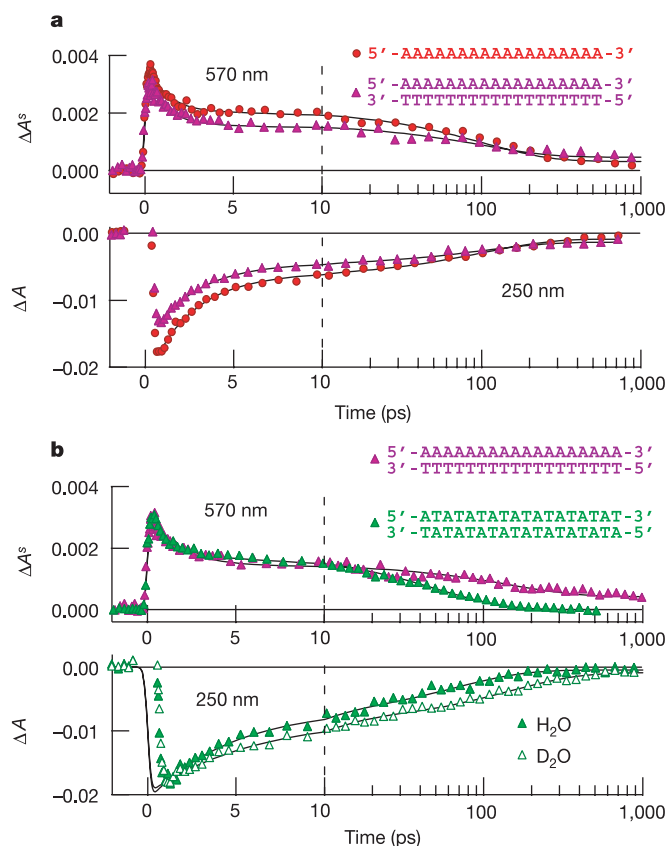


Figure 4 | Singlet excited-state dynamics in double-stranded oligonucleotides. a, Transient absorption at indicated probe wavelengths (top, 570 nm; bottom, 250 nm) following 266 nm excitation of $(dA)_{18}$ ($dT)_{18}$ (triangles). Transients from single-stranded $(dA)_{18}$ (circles) are shown for comparison. All transients can be globally fitted with the same time constants for both oligonucleotides (solid curves, see Supplementary Table 3). Relative amplitudes are significant since all signals were recorded on equal absorbance solutions under identical conditions. **b**, Top, Comparison of transients from the duplexes $(dA)_{18}$ ($dT)_{18}$ (purple triangles) and $d(AT)_9$ ($d(AT)_9$) (green triangles). Smooth curves are from a global fit (see Supplementary Table 4). Signals have been normalized to have the same amplitude near time zero. Bottom panel contrasts the dynamics of $d(AT)_9$ ($d(AT)_9$) in H_2O (filled triangles) and D_2O (open triangles).

tendency to form excimers may explain why most DNA photolesions involve pairs of stacked bases⁵. Less common helical conformations permit stacking by bases on opposite strands, giving rise to inter-strand crosslinks^{26–28}. In B-form DNA, base stacking is limited to bases on the same strand, restricting photoproducts to one strand at a time. Such lesions are readily repaired using the undamaged strand as a template. The ubiquity and versatility of nucleotide excision repair may thus be a consequence of the rapid localization of electronic energy in vertically stacked bases.

METHODS

Oligodeoxyribonucleotides. All samples were obtained from Midland Certified Reagent Company (Midland) as lyophilized powders and dissolved in 25 mM aqueous phosphate buffer containing 0.25 M NaCl. The duplex $(dA)_{18}$ ($dT)_{18}$ was prepared by mixing equimolar amounts of $(dA)_{18}$ and $(dT)_{18}$ stock solutions. Concentrations were determined using extinction coefficients calculated with a nearest-neighbour model²⁹. Both duplex samples, $(dA)_{18}$ ($dT)_{18}$ and $d(AT)_9$ ($d(AT)_9$), were thermally annealed by heating to 70 °C, followed by slow cooling to room temperature. Melting curves and circular dichroism spectra (Supplementary Figs 6 and 7, and Supplementary Methods) confirmed the double-stranded character and B-like conformation of both duplexes. Generally, solutions had equal absorbance at the pump wavelength of 266 nm in order to produce equal numbers of singlet excited states at $t = 0$.

Femtosecond pump–probe measurements. Signals were recorded using Ti:sapphire-based transient absorption spectrometers in Ohio State's Center for Chemical and Biophysical Dynamics. Pump pulses were provided by the third harmonic of the laser fundamental. Probe pulses were obtained at visible wavelengths from a white light continuum, and at UV probe wavelengths from a tunable optical parametric amplifier (Coherent). Pump pulse intensities were typically in the range of $0.2\text{--}2\text{ GW cm}^{-2}$. These low intensities were chosen to minimize the amount of two-photon ionization of the solvent. Probe pulses were always attenuated to have at least ten times less irradiance than the pump pulses. The transmitted probe pulse was detected with an amplified silicon photodiode or a photomultiplier tube after being spectrally filtered by a double grating monochromator or a bandpass filter.

Pump and probe pulses were polarized at the magic angle to eliminate reorientation signals. Transient absorption signals were recorded using a lock-in amplifier referenced to a chopper set to pass every third pulse. The instrument response function was approximately 250 fs or 400 fs for visible or UV probe wavelengths, respectively. Approximately 0.4 ml of the solution under study was held in a 1.2 mm path length cell between calcium fluoride windows. The cell was spun at several hundred r.p.m. to avoid re-excitation of the pumped volume by successive laser pulses. Signals were carefully monitored for photodegradation by UV absorption spectroscopy during the experiments, and solutions were replaced with fresh ones, as needed. In general, little photodegradation was observed during the measurement time with the exception of (dT)₁₈, which degraded rapidly owing to photodimer formation (Supplementary Fig. 5). Control experiments confirmed that photodimer formation has no effect on the signals. All transients at visible probe wavelengths were corrected for two-photon ionization of the solvent as described previously¹³. This correction was not applied at UV wavelengths owing to the small cross-section of the solvated electron below 300 nm. The temperature for all experiments ($T = 28 \pm 2^\circ\text{C}$) was slightly above room temperature owing to weak heating of the sample by the electric motor used to rotate the spin cell.

Received 30 April; accepted 16 June 2005.

- Crespo-Hernández, C. E., Cohen, B., Hare, P. M. & Kohler, B. Ultrafast excited-state dynamics in nucleic acids. *Chem. Rev.* **104**, 1977–2019 (2004).
- Löwdin, P. O. Proton tunneling in DNA and its biological implications. *Rev. Mod. Phys.* **35**, 724–732 (1963).
- Schultz, T. et al. Efficient deactivation of a model base pair via excited-state hydrogen transfer. *Science* **306**, 1765–1768 (2004).
- Abo-Riziq, A. et al. Photochemical selectivity in guanine-cytosine base-pair structures. *Proc. Natl Acad. Sci. USA* **102**, 20–23 (2005).
- Cadet, J. & Vigny, P. in *The Photochemistry of Nucleic Acids* (ed. Morrison, H.) 1–272 (Wiley, New York, 1990).
- Lamola, A. A. & Eisinger, J. Mechanism of thymine photodimerization. *Proc. Natl Acad. Sci. USA* **59**, 46–51 (1968).
- Rahn, R. O. Search for an adenine photoproduct in DNA. *Nucleic Acids Res.* **3**, 879–890 (1976).
- Danilov, V. I., Slyusarchuk, O. N., Alderfer, J. L., Stewart, J. J. P. & Callis, P. R. A theoretical study of the cytosine excimer state: The role of geometry optimization. *Photochem. Photobiol.* **59**, 125–129 (1994).
- Eisinger, J., Guéron, M., Schulman, R. G. & Yamane, T. Excimer fluorescence of dinucleotides, polynucleotides, and DNA. *Proc. Natl Acad. Sci. USA* **55**, 1015–1020 (1966).
- Vigny, P. Fluorescence of polyadenylic acid at room temperature. *C.R. Acad. Sci. D* **277**, 1941–1944 (1973).
- Douhal, A., Kim, S. K. & Zewail, A. H. Femtosecond molecular dynamics of tautomerization in model base pairs. *Nature* **378**, 260–263 (1995).
- Rahn, R. O., Yamane, T., Eisinger, J., Longworth, J. W. & Shulman, R. G. Luminescence and electron spin resonance studies of adenine in various polynucleotides. *J. Chem. Phys.* **45**, 2947–2954 (1966).
- Crespo-Hernández, C. E. & Kohler, B. Influence of secondary structure on electronic energy relaxation in adenine homopolymers. *J. Phys. Chem. B* **108**, 11182–11188 (2004).
- Pecourt, J.-M. L., Peon, J. & Kohler, B. DNA excited-state dynamics: Ultrafast internal conversion and vibrational cooling in a series of nucleosides. *J. Am. Chem. Soc.* **123**, 10370–10378 (2001).
- Dewey, T. G. & Turner, D. H. Laser temperature-jump study of stacking in adenylic acid polymers. *Biochemistry* **18**, 5757–5762 (1979).
- Marguet, S. & Markovitsi, D. Time-resolved study of thymine photodimer formation. *J. Am. Chem. Soc.* **127**, 5780–5781 (2005).
- Hosszu, J. L. & Rahn, R. O. Thymine dimer formation in DNA between 25°C and 100°C . *Biochem. Biophys. Res. Commun.* **29**, 327–330 (1967).
- Ostrowski, T., Maurizot, J.-C., Adeline, M.-T., Fourrey, J.-L. & Clivio, P. Sugar conformational effects on the photochemistry of thymidyl(3'–5')thymidine. *J. Org. Chem.* **68**, 6502–6510 (2003).
- Samoylova, E. et al. Dynamics of photoinduced processes in adenine and thymine base pairs. *J. Am. Chem. Soc.* **127**, 1782–1786 (2005).
- Plützer, C., Hunig, I. & Kleiner, K. Pairing of the nucleobase adenine studied by IR-UV double-resonance spectroscopy and *ab initio* calculations. *Phys. Chem. Chem. Phys.* **5**, 1158–1163 (2003).
- Cukier, R. I. & Nocera, D. G. Proton-coupled electron transfer. *Annu. Rev. Phys. Chem.* **49**, 337–369 (1998).
- Azumi, T., Armstrong, A. T. & McGlynn, S. P. Energy of excimer luminescence. II. Configuration interaction between molecular exciton states and charge resonance states. *J. Chem. Phys.* **41**, 3839–3852 (1964).
- Markovitsi, D., Sharonov, A., Onidas, D. & Gustavsson, T. The effect of molecular organization in DNA oligomers studied by femtosecond fluorescence spectroscopy. *ChemPhysChem* **4**, 303–305 (2003).
- Kelley, S. O. & Barton, J. K. Electron transfer between bases in double helical DNA. *Science* **283**, 375–381 (1999).
- Lewis, F. D., Zuo, X. B., Liu, J. Q., Hayes, R. T. & Wasielewski, M. R. Dynamics of inter- and intrastrand hole transport in DNA hairpins. *J. Am. Chem. Soc.* **124**, 4568–4569 (2002).
- Glisin, V. & Doty, P. The cross-linking of DNA by ultraviolet radiation. *Biochim. Biophys. Acta* **142**, 314–321 (1967).
- Kypr, J., Penazova, H., Sagi, J., Pospisilova, S. & Vorlickova, M. UV light-induced crosslinking of the strands of poly(dA-dT) and related alternating purine-pyrimidine DNAs. *J. Biomol. Struct. Dyn.* **11**, 1225–1236 (1994).
- Douki, T., Laporte, G. & Cadet, J. Inter-strand photoproducts are produced in high yield within A-DNA exposed to UVC radiation. *Nucleic Acids Res.* **31**, 3134–3142 (2003).
- Cantor, C. R., Warshaw, M. M. & Shapiro, H. Oligonucleotide interactions. 3. Circular dichroism studies of the conformation of deoxyoligonucleotides. *Biopolymers* **9**, 1059–1077 (1970).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the NIH and performed using laser instrumentation in Ohio State's Center for Chemical and Biophysical Dynamics funded by the NSF. B.K. acknowledges support from the Alexander von Humboldt Foundation.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to B.K. (kohler@chemistry.ohio-state.edu).

Seasonal oscillations in water exchange between aquifers and the coastal ocean

Holly A. Michael¹, Ann E. Mulligan² & Charles F. Harvey¹

Ground water of both terrestrial and marine origin flows into coastal surface waters as submarine groundwater discharge, and constitutes an important source of nutrients, contaminants and trace elements to the coastal ocean^{1–5}. Large saline discharges have been observed by direct measurements^{3,6–10} and inferred from geochemical tracers^{11–13}, but sufficient seawater inflow has not been observed to balance this outflow. Geochemical tracers also suggest a time lag between changes in submarine groundwater discharge rates^{12,14} and the seasonal oscillations of inland recharge that drive groundwater flow towards the coast. Here we use measurements of hydraulic gradients and offshore fluxes taken at Waquoit Bay, Massachusetts, together with a modelling study of a generalized coastal groundwater system to show that a shift in the freshwater–saltwater interface—controlled by seasonal changes in water table elevation—can explain large saline discharges that lag inland recharge cycles. We find that sea water is drawn into aquifers as the freshwater–saltwater interface moves landward during winter, and discharges back into coastal waters as the interface moves seaward in summer. Our results demonstrate the connection between the seasonal hydrologic cycle inland and the saline groundwater system in coastal aquifers, and suggest a potentially important seasonality in the chemical loading of coastal waters.

Studies that have directly measured submarine groundwater discharge (SGD) with networks of seepage meters show that much of the discharging water has salinity near that of sea water^{6,7,9,10}, yet seawater inflow to coastal aquifers has not been observed in sufficient quantity to explain the large saline outflow. Moore¹² used natural radium as a tracer of groundwater discharge to coastal waters and estimated that SGD was $\sim 100 \text{ m}^3 \text{ d}^{-1}$ per m length of shoreline along the southeastern US coast, equivalent to 40% of river discharge. From consideration of the regional freshwater balance, Moore and Church¹³ conclude that most of this discharge is seawater circulation.

Several studies that infer SGD from natural radium measured in coastal waters reveal a seasonal pattern that is out of phase with the recharge cycle. Radium measurements over several years along the South Atlantic Bight indicate that discharge is larger in the summer than in the winter and spring¹², and monthly groundwater fluxes estimated from radium measurements in Rhode Island exhibit a distinct pattern that also peaks in the summer¹⁴. Recharge is lowest in the summer where these studies were conducted along the eastern US coast because the strong seasonal oscillation of evapotranspiration peaks in the summer, dominating a weaker seasonal signal in precipitation. Along the Ganges Delta, radium transported by SGD is also out of phase with the recharge cycle, although the seasonal pattern is shifted. There, radium flux to ocean water is largest in the winter¹⁵, but recharge is highest in the summer because monsoonal rains dominate evapotranspiration. However, natural tracers

measured in sea water provide little insight into the groundwater dynamics that drive SGD, and most direct SGD measurements by seepage meters or hydraulic gradients have been conducted in temperate regions during the summer only^{7,9,10}.

Both the missing source of saline ground water to coastal aquifers and the seasonal pattern in total SGD can be explained by seasonal exchange of saline water between aquifers and the coastal ocean. In most coastal aquifers, freshwater discharge occurs throughout the year because the water table remains above sea level (Fig. 1). In such aquifers, the well-known Ghyben–Herzberg approximation¹⁶ predicts that the depth to the freshwater–saltwater interface below mean sea level is 40 times the water table elevation above sea level, a factor that results from the density difference between salt water and fresh water. Thus, changes in the water table elevation could theoretically be amplified by a factor of 40 in the interface depth, potentially driving large fluxes of saline water between the subsurface and coastal waters.

The Ghyben–Herzberg relation is only approximate for dynamic groundwater systems because it assumes local hydrostatic conditions and no mixing of salt and fresh water, and the theoretical amplification in the interface movement will decrease as dynamic equilibrium is established. However, it remains true that motion of the water table will drive interface movement. As recharge lifts the water table, the interface moves seaward, fresh water replaces salt water, more fresh water is drawn from inland, and saltwater discharge is driven into the

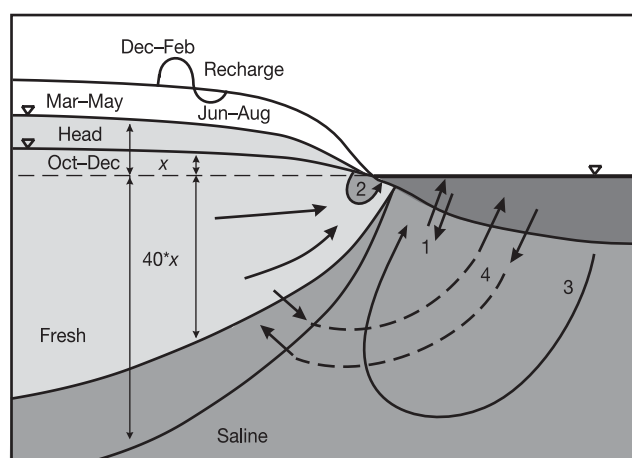


Figure 1 | Saline groundwater circulation in a simple coastal groundwater system with a Ghyben–Herzberg interface. Mechanisms include: (1) tidal pumping, (2) nearshore circulation due to tides and waves, (3) saline circulation driven by dispersive entrainment and brackish discharge, and (4) seasonal exchange.

¹Department of Civil and Environmental Engineering, MIT, Cambridge, Massachusetts 02139, USA. ²Marine Policy Center, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA.

coastal ocean; the opposite set of fluxes occurs when the water table falls. Our numerical simulations of density-coupled groundwater flow and salt transport in unconfined coastal aquifers relax the assumptions of the Ghyben-Herzberg relation, and predict that seasonal recharge oscillations drive a saltwater flux in and out of the aquifer that can be greater than the freshwater discharge (Fig. 2).

Sensitivity analysis confirms that saline discharge lags inland fresh recharge under a variety of hydrogeologic conditions (Supplementary Figs 4 and 5). The hydraulic head in aquifers has been shown to lag recharge in the field by 0–3 months (refs 17, 18) because recharging water takes time to percolate through the unsaturated zone, and because the water table will continue to rise in response to positive recharge past the time of peak recharge. Our numerical simulations of the full dynamic system show that peak saline discharge may lag 1–5 months behind peak recharge (Fig. 2, Supplementary Figs 4 and 5). Seasonal exchange dominates dispersive circulation in every simulation, except when dispersivity and hydraulic conductivity are largest. Thus, simulated saline inflow during the winter can explain both a decrease in total wintertime discharge and the observed net saline discharge in the summer. Seasonal hydrologic cycles occur in many regions of the world¹⁹, and seasonal seawater exchange is predicted for every set of parameters in our simulations of homogeneous unconfined coastal aquifers (Fig. 2 and Supplementary Fig. 2). Furthermore, layered aquifer systems may exhibit enhanced seasonal exchange due to an increase in the length of the freshwater–saltwater interface (Supplementary Fig. 6). Thus, seasonal saline exchange probably exists in a wide range of coastal systems.

To investigate the hypothesis of seasonal saltwater exchange, vertical hydraulic gradients were measured during the winter (February 2004) in the sediments beneath Waquoit Bay, Massachusetts,

when ice cover enabled piezometer installation but prevented the use of seepage meters. Saline discharge (~70% of total discharge in August 2003) has been extensively characterized by direct seepage meter measurement during each summer from 1999 to 2001⁷ and in August 2002 and 2003 (this study) in the same location. The field results (Fig. 3) indicate that the direction of saline flow is reversed between the winter and summer. Net downward hydraulic gradients from the bay into the aquifer, over a complete tidal cycle, were measured at five locations corresponding to the band of highest saline discharge observed during the summer.

Both the fresh and saline discharge observed at Waquoit Bay in the summer are probably a result of flow in the unconfined aquifer, and seasonal interface motion explains the saline component of summer discharge as well as winter inflow. The timing is shown through analysis of meteorological data from Long Pond in Falmouth, Massachusetts²⁰, which indicates positive recharge from late autumn to early spring during the study period (1999–2004), and greater evapotranspiration than precipitation from late spring through to early autumn (Supplementary Fig. 7). Water levels in wells screened at varying depths in the unconfined aquifer within 6.5 km of the study site²¹ over the same period follow a yearly cycle, peaking in April and dropping to a minimum in December. This time lag between the recharge maximum and the hydraulic head maximum is consistent with the numerical results.

Several mechanisms other than seasonal exchange have been hypothesized to drive saltwater circulation^{2,5,10,22–24}, including tides, wave run-up on the beach, and dispersive entrainment of saline water into fresh discharge. These processes will not drive the seasonal shifts from net outflow to inflow that we observe because they produce inflows and outflows that balance over shorter timescales. Furthermore, the circulation produced by these processes (labelled 1, 2 and 3 in Fig. 1) appears to be of secondary importance to seasonal upland forcing at our field site. Tidal pumping (process 1 in Fig. 1) drives water into the aquifer at high tide and out at low tide. Because elastic storage contributes little water over the timescale of tides, significant tidal pumping occurs only near enough to shore such that water table movement can accommodate inflows and outflows. In Waquoit Bay, tidal pumping occurs in a zone less than 15 m from shore, but net inflow has not been observed in this zone over a tidal cycle, and the

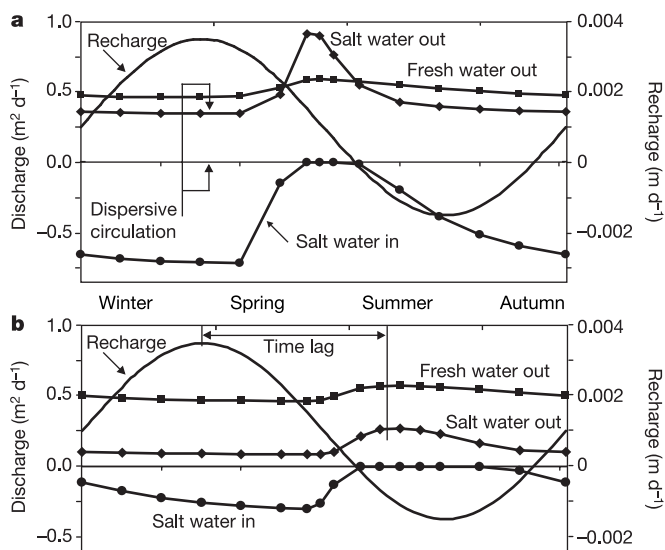


Figure 2 | Simulated total fresh and saline fluxes across the sea floor per metre length of shoreline. Seasons are approximate for a typical yearly recharge cycle within the US. Results are presented over one simulated year for a fixed parameter set (aquifer thickness = 100 m, annual average recharge = 0.001 m d⁻¹, longitudinal dispersivity = 2 m and transverse dispersivity = 0.1 m) and two values of hydraulic conductivity, K : **a**, $K = 5 \times 10^{-4}$ m s⁻¹, and **b**, $K = 1 \times 10^{-4}$ m s⁻¹. Dispersive circulation is apparent as saline outflow during times of net inflow. Additional simulations (see also Supplementary Figs 2–5) indicate that increasing either hydraulic conductivity or aquifer thickness results in an increase in both seasonal outflow and dispersive entrainment but reduces lag, increasing average recharge reduces seasonal outflow compared to fresh outflow and decreases lag, and increasing dispersivity increases dispersive circulation, slightly decreases seasonal outflow, and has no effect on lag. Seasonal changes in freshwater discharge are much less than in saltwater discharge or inland recharge.

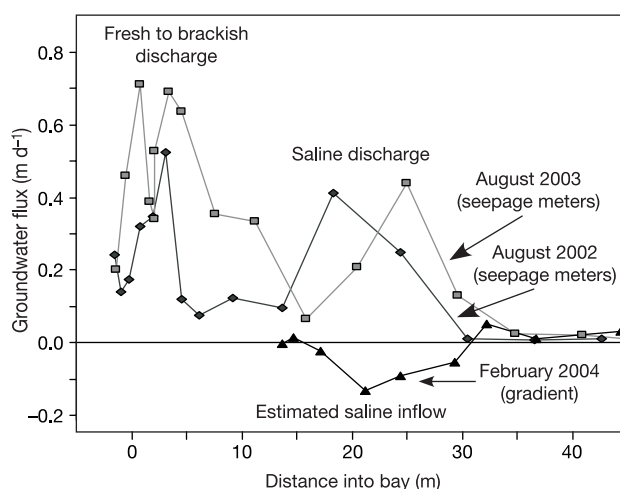


Figure 3 | Submarine groundwater discharge into Waquoit Bay, Massachusetts. August seepage meter measurements reveal primarily saline discharge over a tidal cycle in both 2002 and 2003. In 2003, net saline discharge was 7.5 m³ d⁻¹, and net fresh discharge was 3.5 m³ d⁻¹ per m length of shoreline. February hydraulic gradient measurements indicate saline inflow at the location where peak outflow was measured during summer. An upward hydraulic gradient and fresh pore water were observed beyond 50 m from shore beneath a low-permeability muck cap and are discussed in Supplementary Information (Supplementary Fig. 12).

magnitude of exchange is much less than the saline discharge observed farther from shore.

Wave run-up and tides (process 2 in Fig. 1) create inflow of sea water in the intertidal zone that discharges offshore²³. This circulation was investigated in Waquoit Bay with novel seepage meters that operate in the shallow intertidal zone (Supplementary Fig. 8), and also with a sodium bromide tracer injected near the high tide mark. The measured intertidal inflow was much less than that needed to balance saline outflow, and the tracer test results show that water that enters the sediment at high tide discharges between 2 and 3 m from the position of high tide, much closer to shore than the bulk of saline discharge (Supplementary Fig. 9). Furthermore, bromide tracer was never detected at depths greater than 1.2 m, indicating that saline circulation due to tides and waves is confined to the shallow intertidal zone.

Finally, salt dispersion along the freshwater–saltwater interface can drive large-scale saltwater circulation²² (process 3 in Fig. 1). Kohout²⁴ estimated this flux to be roughly 10% of the seaward flow of ground water at a Florida field site. In Waquoit Bay, however, net inflow seaward of the high discharge zone was observed in only one location during each August experiment, both times within measurement error of zero flow, and most saline water discharges far offshore of the freshwater interface.

Our characterization of the location and magnitude of the four mechanisms in Waquoit Bay (Supplementary Fig. 10) indicates that saline discharge due to the seasonal exchange mechanism ($\sim 3.7 \text{ m}^3 \text{ d}^{-1}$ per m length of shore) is similar to or greater than the total saline discharge due to other circulation mechanisms ($\sim 3.3 \text{ m}^3 \text{ d}^{-1} \text{ m}^{-1}$) during the summer. In summary, three previously known mechanisms for saline circulation result in zero net saline outflow over a tidal cycle and none can explain the large net saline discharge observed in Waquoit Bay during the summer, leaving seasonally-induced saline cycling as the likely explanation for our observations.

Seasonal exchange from coastal aquifers may account for a large component of SGD where recharge cycles are significant, whether natural or anthropogenic in origin. The global extent of seasonal saline exchange is currently unknown, but yearly recharge cycles appear to be widespread, suggesting that seasonal interface movement could be a potentially important driver of saline water exchange along many coastlines. SGD is known to transport nutrients and contaminants that affect coastal ecosystems^{1,25}, and the solutes delivered by saline discharge can be as important (or more important) as those delivered by fresh SGD^{3,6}. Seasonal exchange may affect sediment chemistry by subjecting coastal sediments to flushing by typically oxic sea water from above during inflow periods, and flushing with often anoxic water from below during discharge periods. In areas along the eastern US coast, the greatest saltwater discharge may occur in the summer when biological activity is maximum and river flow is minimum, so input of nutrients may be of particular importance. Furthermore, the chemistry of saltwater discharge may vary seasonally because the first saline water to discharge has most recently entered the aquifer, and the last saline water to discharge in the yearly cycle has had the longest subsurface residence time. This work demonstrates the connection between the inland seasonal hydrologic cycle and the saline groundwater system in coastal aquifers, proposes a mechanism for saltwater circulation within aquifers that results in net saline outflow during several months each year, and suggests the potential for seasonality in chemical loading of coastal waters.

METHODS

Numerical simulations. Two-dimensional variable-density simulations were performed using the finite element model FEFLOW²⁶. The simulated domain extended 500 m landward and 200 m seaward from the shoreline for aquifers 20 m and 100 m thick (Supplementary Fig. 1), and the number of elements ranged from 93,532 to 597,638. The aquifer was simulated as unconfined and

spatially homogeneous, with zero flow and zero mass transport boundary conditions along the base and sides. The recharge boundary condition along the landward model top varied sinusoidally in time, with an average value of 0.002 or 0.001 m d^{-1} and amplitude of 0.0025 m d^{-1} . The sea-floor boundary was a constant head with constant concentration where flow was inward, zero concentration gradient where flow was outward. Eight simulations were run: two are shown in Fig. 2, and six are presented in Supplementary Information (Supplementary Table 1, Supplementary Figs 1–5), with varying hydraulic conductivity, dispersivity, aquifer thickness and average recharge.

Seepage meters. During August 2002 and 2003, 20 conventional seepage meters²⁷, enough to overcome local variability⁷, were installed in Waquoit Bay. The meters were aligned in two transects, 1 m apart, and extended 50 m into Waquoit Bay perpendicular to the shoreline in the same location where 40 seepage meters were used in 1999 and 2000⁷. Eight novel intertidal seepage meters (Supplementary Fig. 8) were placed in the nearshore zone of this transect, the locations varying with the position of the tide. Unlike conventional seepage meters, these intertidal meters are not submerged, enabling measurement of groundwater inflow and outflow in very shallow water depths. Seepage meter bags were pre-filled with sufficient bay water to allow detection of either inflow or outflow, and collected at least every two hours over one tidal cycle. Discharge salinity was calculated from conductivity probe measurements in samples from the conventional seepage meter bags before and after deployment, and by refractometer measurements from porewater samples 3 cm below the sediment surface next to each intertidal seepage meter. Groundwater flux and salinity measurements agree well between adjacent conventional and intertidal seepage meters.

Piezometers. In February 2004, 11 piezometers were placed through bay ice along the summer seepage meter transect to depths of 0.6–0.9 m. The water levels in each piezometer and the bay were measured with an electronic water level meter approximately every 1–2 h during daylight hours. The salinity of piezometer samples and baywater profiles were measured every few hours with a conductivity probe. The average vertical hydraulic gradient over a tidal cycle from two days of data with similar tidal variation was calculated by correcting for density differences in the bay and piezometer water columns. Slug test data along this transect were converted to estimates of hydraulic conductivity (Supplementary Fig. 11). The density-corrected hydraulic gradients and interpolated hydraulic conductivities were converted to flux estimates by Darcy's law.

Measured seepage rates, both inflow and outflow, from seepage meters correlate closely to hydraulic gradients measured in adjacent piezometers when tested concurrently, so these summer seepage meter measurements and winter piezometer results can be compared directly.

Tracer test. On 27 August 2001, a sodium bromide solution was injected into the beach at the head of Waquoit Bay near the high tide mark, during high tide. The 0.243 M injection solution, with the density of seawater, 1.025 kg l^{-1} , was used to track the movement of the infiltrating bay water. Twenty-one piezometers were driven to depths of 0.3–1.4 m, nested in groups placed approximately every 0.6 m to a distance 4.3 m bayward from the injection point. Small-volume porewater samples were extracted every 1–2 h during daylight for four consecutive days, 32 sample times in all. A conductivity probe and a multi-meter connected to a bromide electrode were used to measure the total conductivity (in mS cm^{-1}) and Br^- concentration (in mV calibrated to mol l^{-1}) of each sample.

Received 25 March; accepted 9 August 2005.

- Johannes, R. E. The ecological significance of the submarine discharge of groundwater. *Mar. Ecol. Prog. Ser.* **3**, 365–373 (1980).
- Moore, W. S. The subterranean estuary: a reaction zone of ground water and sea water. *Mar. Chem.* **65**, 111–125 (1999).
- Simmons, G. M. Importance of submarine groundwater discharge (SGWD) and seawater cycling to material flux across sediment water interfaces in marine environments. *Mar. Ecol. Prog. Ser.* **84**, 173–184 (1992).
- Slomp, C. P. & Van Cappellen, P. Nutrient inputs to the coastal ocean through submarine groundwater discharge: controls and potential impact. *J. Hydrol.* **295**, 64–86 (2004).
- Burnett, W. C., Bokuniewicz, H., Huettel, M., Moore, W. S. & Taniguchi, M. Groundwater and pore water inputs to the coastal zone. *Biogeochemistry* **66**, 3–33 (2003).
- Kim, G., Lee, K. K., Park, K. S., Hwang, D. W. & Yang, H. S. Large submarine groundwater discharge (SGD) from a volcanic island. *Geophys. Res. Lett.* **30** doi:10.1029/2003GL018378 (2003).
- Michael, H. A., Lubetsky, J. S. & Harvey, C. F. Characterizing submarine groundwater discharge: a seepage meter study in Waquoit Bay, Massachusetts. *Geophys. Res. Lett.* **30** doi:10.1029/GL016000 (2003).
- Robinson, M., Gallagher, D. & Reay, W. Field observations of tidal and seasonal variations in ground water discharge to tidal estuarine surface water. *Ground Wat. Monit. Remediat.* **18**, 83–92 (1998).

9. Smith, L. & Zawadzki, W. A hydrogeologic model of submarine groundwater discharge: Florida intercomparison experiment. *Biogeochemistry* **66**, 95–110 (2003).
10. Taniguchi, M., Ishitobi, T. & Saeki, K. Evaluation of time-space distributions of submarine ground water discharge. *Ground Wat.* **43**, 336–342 (2005).
11. Crotwell, A. M. & Moore, W. S. Nutrient and radium fluxes from submarine groundwater discharge to Port Royal Sound, South Carolina. *Aquat. Geochem.* **9**, 191–208 (2003).
12. Moore, W. S. Large groundwater inputs to coastal waters revealed by Ra-226 enrichments. *Nature* **380**, 612–614 (1996).
13. Moore, W. S. & Church, T. M. Submarine groundwater discharge — Reply. *Nature* **382**, 122–122 (1996).
14. Kelly, R. P. & Moran, S. B. Seasonal changes in groundwater input to a well-mixed estuary estimated using radium isotopes and implications for coastal nutrient budgets. *Limnol. Oceanogr.* **47**, 1796–1807 (2002).
15. Moore, W. S. High fluxes of radium and barium from the mouth of the Ganges-Brahmaputra river during low river discharge suggest a large groundwater source. *Earth Planet. Sci. Lett.* **150**, 141–150 (1997).
16. Bear, J., Cheng, A. H.-D., Sorek, S., Ouazar, D. & Herrera, I. in *Seawater Intrusion in Coastal Aquifers — Concepts, Methods and Practices* (ed. Bear, J.) (Kluwer Academic, Dordrecht, 1999).
17. Eltahir, E. A. B. & Yeh, P. A. J. F. On the asymmetric response of aquifer water level to floods and droughts in Illinois. *Wat. Resour. Res.* **35**, 1199–1217 (1999).
18. Changnon, S. A., Huff, F. A. & Hsu, C.-F. Relations between precipitation and shallow groundwater in Illinois. *J. Clim.* **1**, 1239–1250 (1988).
19. Tapley, B. D., Bettadpur, S., Ries, J. C., Thompson, P. F. & Watkins, M. M. GRACE measurements of mass variability in the Earth system. *Science* **305**, 503–505 (2004).
20. Payne, R. Falmouth monthly climate reports. (<http://cis.whoi.edu/science/PO/climate/index.cfm>) (accessed 30 July 2004).
21. US Geological Survey National Water Information System (NWISWeb). (<http://nwis.waterdata.usgs.gov/usa/nwis/gwlevels>) (accessed 30 July 2004).
22. Cooper, H. H. A hypothesis concerning the dynamic balance of fresh water and salt water in a coastal aquifer. *J. Geophys. Res.* **64**, 461–467 (1959).
23. Li, L., Barry, D. A., Stagnitti, F. & Parlange, J. Y. Submarine groundwater discharge and associated chemical input to a coastal sea. *Wat. Resour. Res.* **35**, 3253–3259 (1999).
24. Kohout, F. Cyclic flow of salt water in the Biscayne aquifer of southeastern Florida. *J. Geophys. Res.* **65**, 2133–2141 (1960).
25. Valiela, I. *et al.* Transport of groundwater-borne nutrients from watersheds and their effects on coastal waters. *Biogeochemistry* **10**, 177–197 (1990).
26. Diersch, H. J. G. Interactive, graphics-based finite-element simulation system FEFLOW for modeling groundwater flow, contaminant mass, and heat transport processes. (WASY Ltd., Berlin, Germany, 2002).
27. Lee, D. R. Device for measuring seepage flux in lakes and estuaries. *Limnol. Oceanogr.* **22**, 140–147 (1977).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the WBNERR staff and the many MIT and WHOI faculty and students, USGS personnel, and others who assisted in the field. This work was supported by a graduate research fellowship from the US National Science Foundation.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.F.H. (charvey@mit.edu).

Frozen magma lenses below the oceanic crust

Mladen R. Nedimović¹, Suzanne M. Carbotte¹, Alistair J. Harding², Robert S. Detrick³, J. Pablo Canales³, John B. Diebold¹, Graham M. Kent², Michael Tischer¹ & Jeffrey M. Babcock²

The Earth's oceanic crust crystallizes from magmatic systems generated at mid-ocean ridges. Whereas a single magma body residing within the mid-crust is thought to be responsible for the generation of the upper oceanic crust, it remains unclear if the lower crust is formed from the same magma body, or if it mainly crystallizes from magma lenses located at the base of the crust^{1–3}. Thermal modelling^{4–6}, tomography⁷, compliance⁸ and wide-angle seismic studies⁹, supported by geological evidence^{3,10–18}, suggest the presence of gabbroic-melt accumulations within the Moho transition zone in the vicinity of fast- to intermediate-spreading centres. Until now, however, no reflection images have been obtained of such a structure within the Moho transition zone. Here we show images of groups of Moho transition zone reflection events that resulted from the analysis of ~1,500 km of multi-channel seismic data collected across the intermediate-spreading-rate¹⁹ Juan de Fuca ridge. From our observations we suggest that gabbro lenses and melt accumulations embedded within dunite or residual mantle peridotite are the most probable cause for the observed reflectivity, thus providing support for the hypothesis that the crust is generated from multiple magma bodies.

The Moho transition zone (MTZ), located at the crust–mantle boundary, separates layered gabbros of the crust (derived by magma crystallization) from residual peridotites (generally harzburgites) representing mantle rocks^{10,12}. Mapping of the Oman and the Bay of Islands ophiolite complexes, both of which are inferred to be composed of obducted oceanic lithosphere formed at intermediate- to fast-spreading ridges, has shown that the MTZ is composed of sills and lenses of gabbro intruded into dunite or, occasionally, into harzburgite^{3,10–18}. The thickness of the MTZ can vary from a few metres to over two kilometres^{10,12}. A thin MTZ (<~100 m)¹⁷ is widespread within the mapped ophiolites and is characterized by intense deformation resulting from solid-state flow away from the ridge axis that transposes all lithologic units into parallelism sub-horizontal to the Moho. In the few areas where a localized, thick MTZ (>~100 m)¹⁷ is observed, individual gabbro sills and lenses can reach thicknesses of a few hundred metres¹². Locally steep orientation of foliation planes and lineations within high-temperature peridotite led to the association of thick MTZs of the Oman ophiolite with preserved ancient mantle diapirs, many of which are centred along inferred palaeo-ridge axes¹⁶. Gabbro sills of the thick MTZ display strong magmatic flow structure, with lineations and foliations parallel to those within the surrounding peridotites where solid-state conditions prevailed. These structures were formed during the horizontal flow, carrying upper mantle formations away from the ridge axis¹³.

In our Juan de Fuca reflection sections, the Moho discontinuity is imaged along more than 60% of the survey track (Fig. 1). Because the inferred crustal thickness is remarkably uniform, as inferred from two-way travel times ($2,080 \pm 100$ ms), the stratigraphic level of

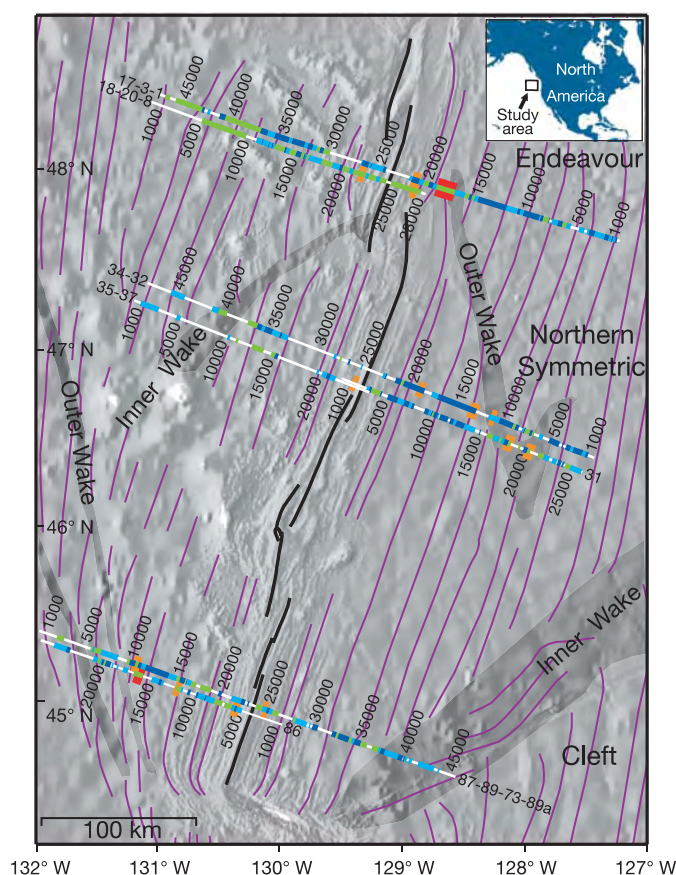


Figure 1 | Study area and strength of the Moho reflection event for the 2002 Juan de Fuca ridge flank seismic profiles is plotted in colour over a sun-illuminated grey bathymetry map. The following reflection strength colour code was used: strong Moho reflection (dark blue); moderate Moho reflection (bright blue); weak Moho reflection (green); no Moho reflection (white). From north to south, seismic profiles cross Endeavour, Northern Symmetric and Cleft ridge segments. Every 5,000th common midpoint (CMP), corresponding to a distance of 31.25 km, is annotated along each profile. Thick red and orange line segments mark sections characterized by sub-Moho reflection events. The MTZ reflection events marked in red were imaged at a high signal-to-noise ratio. The interpreted traces of the ridge axis are indicated with thick black lines and magnetic isochrons¹⁹ with thin purple lines. Overlay areas approximately outline the location of inner and outer propagator wakes. The inset shows the location of the study area with respect to North America.

¹Lamont-Doherty Earth Observatory of Columbia University, 61 Route 9W, PO Box 1000, Palisades, New York 10964-8000, USA. ²Scripps Institution of Oceanography, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0225, USA. ³Woods Hole Oceanographic Institution, 360 Woods Hole Rd, Woods Hole, Massachusetts 02543-1542, USA.

events originating within the MTZ is well constrained. The Moho discontinuity is mostly represented by a single reflection event of variable strength, at places remarkably sharp and strong (Fig. 2a). Seismic modelling studies^{11,20} suggest that areas with a single strong or moderately strong Moho reflection event are probably characterized by an abrupt passage from layered gabbros of the lower crust, to residual peridotites of the uppermost mantle. Areas with weak or no Moho reflection event may be indicative of rough Moho topography that scatters acoustic energy, or may represent thick MTZs composed of thin alternating gabbro and dunite sills, where the ratio of dunite to gabbro gradually increases with depth resulting in a gradual downward velocity increase (Supplementary Discussion 1).

In Fig. 2b, we show the largest and the most prominent group of MTZ reflection events recorded (Supplementary Discussion 2). These sub-horizontal reflections are located on line 17-3-1, more than 30 km east of the ridge axis (Fig. 1). The Moho reflection event in this section of line 17-3-1 is weak but traceable, placing the recorded events just below the crust. For the purposes of discussion and following ref. 12, we assume that the geophysical Moho and petrological Moho are equivalent, and that the MTZ is therefore the uppermost part of the mantle. The spatial character of the imaged

subcrustal reflection events is in excellent agreement with the geometry of the thick MTZ as described by ophiolite studies¹². The anomalous MTZ area is approximately 10 km wide, with individual bright events showing 4–5 km of lateral continuity and a combined thickness of over 2,000 m in the central region, comparable to the maximum thickness mapped in ophiolites^{10,12}.

High signal-to-noise ratio of the imaged MTZ events presented in the migrated stack of Fig. 2b indicates that these reflections probably arise at sharp boundaries between layers of significantly different physical properties. The strength of the observed sub-Moho events may also be partially due to constructive interference from the top and the bottom sill reflections. Assuming that the MTZ events are caused by gabbro sills with a P-wave velocity of 7 km s^{-1} , and that the dominant data signal frequency is $\sim 10\text{--}15 \text{ Hz}$, the strongest responses will be generated by gabbro sills with thicknesses from $\sim 50\text{--}150 \text{ m}$, and will be characterized by single reflection events. This is in agreement with the thickness of large gabbro sills mapped in ophiolites. Sills with thicknesses $< \sim 20 \text{ m}$ cannot be imaged because of destructive wave interference. Images of sills with a thickness $> \sim 200 \text{ m}$ will be characterized by paired top and bottom reflection events of opposite polarity, a pattern not observed in our seismic sections.

Insight about the lithologies causing the reflections can in principle be obtained by analysis of reflection amplitude variation as a function of offset or angle of incidence (AVO/AVA). However, the signal-to-noise ratio of our prestack data was much too low for the standard AVO analysis. We therefore computed trace envelopes (Fig. 3a) for a partially stacked super CMP gather positioned over the brightest group of subcrustal reflections from Fig. 2b, and compared them with calculated amplitude curves for rays reflected from the potential rock interfaces of a thick MTZ (Fig. 3b–f) (Supplementary Discussion 3). The observed reflection strength of the MTZ events seems to gradually weaken with the increasing source-receiver offset, with somewhat greater amplitude drop at offsets larger than about 4 km. Based on the relationship between modelled and observed reflection strength curves (red and blue lines in Fig. 3b and c), we speculate that the imaged reflections were generated at contacts between solid gabbro sills and host dunite, a structural relationship frequently observed in the thick MTZs of the Oman ophiolite¹³. We cannot resolve (see Fig. 3) whether the host rock for the gabbro sills is dunite or residual mantle peridotite³.

The possibility that the strongest MTZ reflections in Fig. 3a were generated at the contact between ultramafic host rocks and gabbro-melt, however, cannot be eliminated. Nevertheless, such an interface is less likely because the ridge axis is located at a significant distance westward ($\sim 33 \text{ km}$), and the predicted reflection amplitude fall-off with offset for a dunite–gabbro–melt interface seems more rapid than that observed (Fig. 3b). Moreover, the reflection amplitudes for the dunite–gabbro–melt contact (Fig. 3b) are much larger than for the corresponding solid–solid interface (Fig. 3c), but high signal-to-noise ratio subcrustal reflections are not identified in the prestack data. The graph shown in Fig. 3f indicates that the lower boundary of thick MTZs, usually represented by a transition from dunite to harzburgite, is transparent for reflection imaging.

In addition to the events shown in Fig. 2b, we were able to identify a number of other areas with complex Moho and sub-Moho reflections along our survey track (Supplementary Discussion 4). The distribution of all imaged subcrustal reflection areas extending for at least 3 km is presented in Fig. 1. The identified thick MTZs form two distinct groups: those centred near the ridge axis and those found beyond ($> \sim 20 \text{ km}$), such as the one shown in Fig. 2b. The thick MTZs found near the ridge axis seem to be uniformly and symmetrically distributed across the ridge axis. An example of a series of subcrustal reflection events located less than 10 km from the ridge axis on line 87-89-73-89a is shown in Fig. 4. Although our data do not provide direct constraints on the lithologic nature of the subcrustal interfaces imaged in the vicinity of the ridge axis, ophiolite

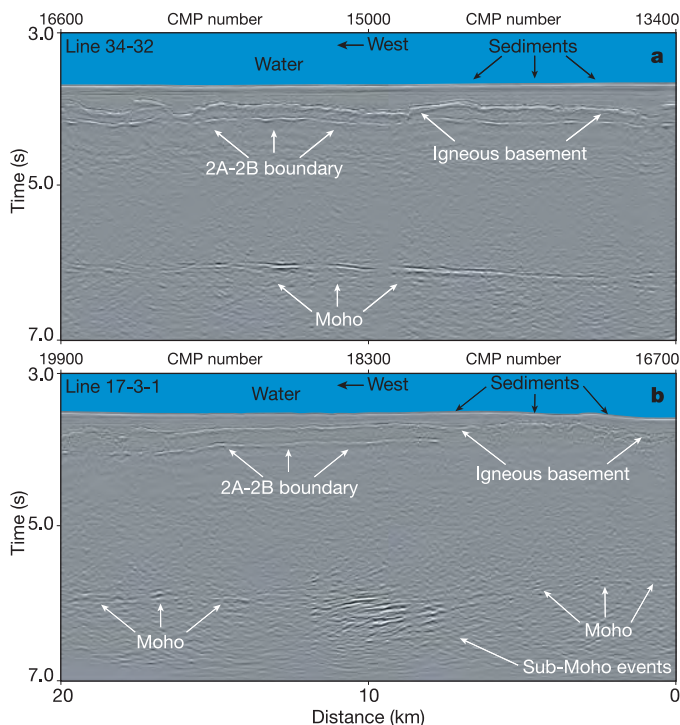


Figure 2 | Seismic reflection structure of the oceanic crust characterized by thick and thin MTZs. **a, b,** Shown are two 20 km-long sections, one of line 34-32 reflection image (**a**) and the other of line 17-3-1 reflection image (**b**), using greyscale variable density plots. The distance to depth ratio is approximately 1:1. Oceanic crustal structure in both sections is imaged remarkably well. Panel **a** is characterized by a mostly continuous and sharp image of the igneous basement surface, the layer 2A/2B boundary and the Moho discontinuity. Panel **b** shows all the same events as panel **a** but differs from it because it also includes a set of sub-Moho reflections, and because it is characterized by a weaker Moho reflection and a layer 2A/2B boundary event that is not imaged across the whole section. Weak events found a few hundred milliseconds below the layer 2A/2B boundary, that may appear to be related to the boundary between the layer 2B and layer 3, are processing artefacts caused by merging of sections with a different content of signal frequencies. The upper parts of the sections include all recorded signal frequencies (2–7–100–125 Hz). To improve the Moho reflection event, only the low signal frequencies (2–7–20–40 Hz) are kept for the deeper parts of the sections. For details about the data acquisition and processing, see Supplementary Methods 1 and 2.

studies^{3,15,17} and other geophysical investigations^{7–9} indicate that gabbro-melt lenses are emplaced in the MTZ near the ridge axis. Therefore, the observed subcrustal reflections, such as those presented in Fig. 4, are more likely to be generated at the interface between ultramafic rocks and gabbro-melt than the subcrustal reflections located farther away from the ridge axis.

Away from the ridge axis ($> \sim 20$ km), the instances of thick MTZs seem to correlate with the location of outer propagator wakes defined by magnetic isochrons (Fig. 1). The general absence of detectable thick MTZs away from the ridge axis, and outside the outer propagator wake area, suggests a highly dynamic environment at the crust–mantle boundary up to a distance of a few tens of kilometres from the spreading centre. Within this area, MTZ gabbro sills and dunite layers both form and can cease to exist. Structural mapping of the Oman ophiolite provides evidence for two possible mechanisms responsible for thinning of the thick MTZs: tectonic stretching and upward magma discharge^{3,17}.

The thick MTZs found $> \sim 20$ km away from the ridge axis, and in particular the largest ones marked in red in Fig. 1, are distinct from the near ridge axis features in size, shape and thickness. Based on their spatial association with the outer propagator wakes, we suggest that these preserved thick MTZs were formed by intrusion of melt into the crust–mantle boundary area within the zone of active

spreading on the propagating ridge segment. At the propagator tips, the spreading breaks into a relatively cool lithosphere that might allow for large magma bodies to be emplaced and solidified. Ophiolite mapping and our data constrain the maximum thickness of emplaced magma bodies to ~ 150 – 200 m, and their maximum diameter to several kilometres. Existence of isolated magma bodies at propagator tips, which appear to have experienced rapid cooling and high fractional crystallization, is supported by the common occurrence of high amplitude magnetic anomalies and high Fe-Ti basalts^{21,22}. Large intrusions of basaltic melt into previously accreted, and therefore older and colder, oceanic lithosphere are also documented in the Maqсад diapir area of the Oman ophiolite^{23,24}, and were associated with the opening of a propagator¹⁸. Two of the imaged thick MTZs (marked red in Fig. 1), characterized by strong reflection events that we believe are indicative of the thick (~ 50 – 150 m) gabbro sills, might also be associated with diapirism linked to opening of the propagators. In both cases the imaged sills appear to have a ridge-ward dip and to be located within the newly accreted crust, which itself is characterized by smoothest topography suggesting locally abundant melt supply. Unlike the crust forming the inside wakes, which is rotated and sheared²¹, the crust forming the outer propagator wakes experiences little deformation²² and therefore provides a sheltered environment in which thick MTZs may be preserved. Interestingly, of the three MTZ melt lenses identified along the East Pacific Rise 11° – 13° N area using PmS waves⁹, two are associated with the outer wakes of migrating overlapping spreading centres.

Our evidence suggests that sill emplacement in the MTZ may be a common feature beneath intermediate spreading centres but the sills themselves are short-lived. Only sills within the thick MTZs formed at propagator tips seem to remain preserved after being accreted to outer propagator wakes. The melt lenses forming these thick MTZs

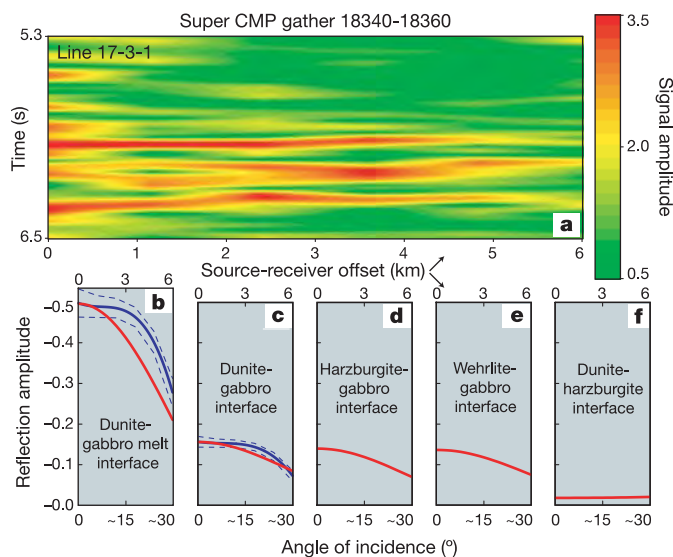


Figure 3 | Modelled and measured reflection amplitude versus offset dependence for the MTZ events shown in Fig. 2b. **a**, Trace envelope of the partially stacked super CMP gather 18340–18360. The super gather is composed of 21 adjacent CMP gathers approximately centred at the middle of the group of bright subcrustal reflection events shown in Fig. 2b. Gather data traces were amplitude corrected only for the geometrical propagation effects and shot and receiver ‘surface’ inconsistencies, sorted to 1-km-wide partial source-receiver offset gathers and stacked. To estimate the signal reflection strength as a function of arrival time, the partial stacked super gather data were then transformed to trace envelope and plotted in colour. **b–f**, Ray amplitude, in red, as a function of the source-receiver offset for the potential rock interfaces of the thick MTZ shown in Fig. 2b (see Supplementary Methods 3). Lithologic interfaces were formed by superimposing gabbro ($v_p = 7,000 \text{ m s}^{-1}$; $v_s = 3,750 \text{ m s}^{-1}$; $\rho = 2,900 \text{ kg m}^{-3}$), dunite ($v_p = 8,450 \text{ m s}^{-1}$; $v_s = 4,850 \text{ m s}^{-1}$; $\rho = 3,300 \text{ kg m}^{-3}$), harzburgite ($v_p = 8,300 \text{ m s}^{-1}$; $v_s = 4,850 \text{ m s}^{-1}$; $\rho = 3,300 \text{ kg m}^{-3}$), wehrlite ($v_p = 8,100 \text{ m s}^{-1}$; $v_s = 4,650 \text{ m s}^{-1}$; $\rho = 3,300 \text{ kg m}^{-3}$) and gabbro melt ($v_p = 3,200 \text{ m s}^{-1}$; $v_s = 0 \text{ m s}^{-1}$; $\rho = 2,900 \text{ kg m}^{-3}$)^{10,25}. For the inverted interfaces of **b–f**, only the sign (polarity) of the reflection amplitudes changes. Solid and dashed blue lines in **b** and **c** are the cubic fit to the average maximum amplitude for the events shown in **a** and the computed error bounds, respectively. The observed amplitude decay curve is scaled so that it matches **b** and **c** modelled amplitudes at zero-offset.

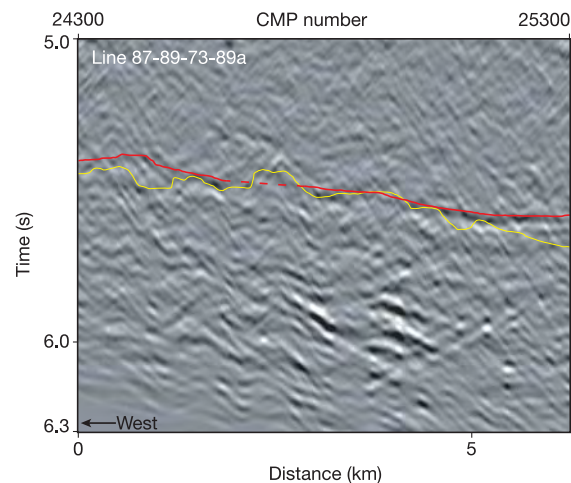


Figure 4 | A series of subcrustal reflection events recorded on line 87-89-73-89a is displayed at the distance to depth ratio of about 1:1. The reflection events are located approximately 7 to 8 km east of the ridge axis (see Fig. 1) and a few hundred milliseconds below the Moho reflection event, which in this section of line 87-89-73-89a is of moderate strength. Our interpretation of the Moho discontinuity location is shown with a continuous red line; dashed where inferred. The yellow line is the igneous basement pick delayed by 2,300 ms, the average two-way travel time through the crust in the vicinity of the Cleft ridge. The presented events appear to exhibit a very mild dip away from the axis but the form and strength of these sub-Moho reflections might be imaged inaccurately due to focusing and defocusing of the acoustic energy in the areas close to the ridge axis, where the seafloor topography is generally rougher. Nevertheless, the recorded signals are probably true sub-Moho reflections as we generally do not observe scattered energy at earlier times, and cannot find an explanation for its focusing just below the Moho discontinuity.

are thought to have experienced repetitious magma expulsion with continuous melt replenishment^{3,15}. Hence, the interpreted large gabbro-melt sills within the thick MTZ shown in Fig. 2b may represent the first images of frozen subcrustal magma chambers.

Received 13 January; accepted 15 June 2005.

- Sinton, J. M. & Detrick, R. S. Mid-ocean ridge magma chambers. *J. Geophys. Res.* **97**, 197–216 (1992).
- Phipps Morgan, J. & Chen, Y. J. The genesis of oceanic crust; magma injection, hydrothermal circulation, and crustal flow. *J. Geophys. Res.* **98**, 6283–6297 (1993).
- Kelemen, P. B. & Aharonov, E. in *Faulting and Magmatism at Mid-Ocean Ridges* (eds Buck, W. R., Delaney, P. T., Karson, J. A. & Lagabriele, Y.) 267–289 (Geophysical Monograph 106, AGU, Washington DC, 1998).
- Garrido, C. J., Kelemen, P. B. & Hirth, G. Variation of cooling rate with depth in lower crust formed at an oceanic spreading ridge: Plagioclase crystal size distributions in gabbros from the Oman ophiolite. *Geochem. Geophys. Geosyst.* **2**, doi:10.1029/2000GC000136 (2001).
- Cherkaoui, A. S. M., Wilcock, W. S. D., Dunn, R. A. & Toomey, D. R. A numerical model of hydrothermal cooling and crustal accretion at a fast spreading mid-ocean ridge. *Geochem. Geophys. Geosyst.* **4**, doi:10.1029/2001GC000215 (2003).
- MacLennan, J., Hulme, T. & Singh, S. C. Thermal models of oceanic crustal accretion: Linking geophysical, geological and petrological observations. *Geochem. Geophys. Geosyst.* **5**, doi:10.1029/2003GC000605 (2004).
- Dunn, R. A., Toomey, D. R., Detrick, R. S. & Wilcock, W. S. D. Continuous mantle melt supply beneath an overlapping spreading center on the East Pacific Rise. *Science* **291**, 1955–1958 (2001).
- Crawford, W. C. & Webb, S. C. Variations in the distribution of magma in the lower crust and at the Moho beneath the East Pacific Rise at 9°–10° N. *Earth Planet. Sci. Lett.* **203**, 117–130 (2002).
- Garmany, J. Accumulations of melt at the base of young oceanic crust. *Nature* **340**, 628–632 (1989).
- Karson, J. A., Collins, J. A. & Casey, J. F. Geologic and seismic velocity structure of the crust/ mantle transition in the Bay of Islands ophiolite complex. *J. Geophys. Res.* **89**, 6126–6138 (1984).
- Collins, J. A., Brocher, T. M. & Karson, J. A. Two-dimensional seismic reflection modeling of the inferred fossil oceanic crust/mantle transition in the Bay of Islands Ophiolite. *J. Geophys. Res.* **91**, 12520–12538 (1986).
- Benn, K., Nicolas, A. & Reuber, I. Mantle-crust transition zone and origin of wehrlitic magmas; evidence from the Oman Ophiolite. *Tectonophysics* **151**, 75–85 (1988).
- Boudier, F. & Nicolas, A. Nature of the Moho transition zone in the Oman ophiolite. *J. Petrol.* **36**, 777–796 (1995).
- Boudier, F., Nicolas, A. & Ildefonse, B. Magma chambers in the Oman Ophiolite; fed from the top and the bottom. *Earth Planet. Sci. Lett.* **144**, 239–250 (1996).
- Korenaga, J. & Kelemen, P. B. Origin of gabbro sills in the Moho transition zone of the Oman Ophiolite; implications for magma transport in the oceanic lower crust. *J. Geophys. Res.* **102**, 27729–27749 (1997).
- Nicolas, A., Boudier, F., Ildefonse, B. & Ball, E. Accretion of Oman and United Arab Emirates ophiolite; discussion of a new structural map. *Mar. Geophys. Res.* **21**, 147–179 (2000).
- Jousselin, D. & Nicolas, A. The Moho transition zone in the Oman Ophiolite; relation with wehrlites in the crust and dunites in the mantle. *Mar. Geophys. Res.* **21**, 229–241 (2000).
- Godard, M., Jousselin, D. & Bodinier, J.-L. Relationships between geochemistry and structure beneath a paleo-spreading centre; a study of the mantle section in the Oman Ophiolite. *Earth Planet. Sci. Lett.* **180**, 133–148 (2000).
- Wilson, D. S. Confidence intervals for motion and deformation of the Juan de Fuca Plate. *J. Geophys. Res.* **98**, 16053–16071 (1993).
- Brocher, T. M., Karson, J. A. & Collins, J. A. Seismic stratigraphy of the oceanic Moho based on ophiolite models. *Geology* **13**, 62–65 (1985).
- Hey, R., Duenebier, F. K. & Morgan, W. J. Propagating rifts on midocean ridges. *J. Geophys. Res.* **85**, 3647–3658 (1980).
- Hey, R., Kleinrock, M. C., Miller, S. P., Atwater, T. M. & Searle, R. C. Sea Beam/ deep-tow investigation of an active oceanic propagating rift system, Galapagos 95.5 W. *J. Geophys. Res.* **91**, 3369–3393 (1986).
- Ceuleneer, G., Monnereau, M. & Amri, I. Thermal structure of a fossil mantle diapir inferred from the distribution of mafic cumulates. *Nature* **379**, 149–153 (1996).
- Amri, I., Benoit, M. & Ceuleneer, G. Tectonic setting for the genesis of oceanic plagiogranites; evidence from a paleo-spreading structure in the Oman Ophiolite. *Earth Planet. Sci. Lett.* **139**, 177–194 (1996).
- Carlson, R. L. in *Handbook of Elastic Properties, Liquids, and Gases* (eds Levy, M., Bass, H. & Stern, R.) 377–461 (Academic Press, New York, 2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to P. B. Kelemen and W. R. Buck for their reviews. K. Vasudevan provided us with a stacked section of the Canadian line 1989–15, shown in Supplementary Information. The Doherty and National Science Foundations supported this work.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.R.N. (mladen@ldeo.columbia.edu).

Methanotrophic symbionts provide carbon for photosynthesis in peat bogs

Ashna A. Raghoebarsing¹, Alfons J. P. Smolders², Markus C. Schmid¹, W. Irene C. Rijpstra⁴, Mieke Wolters-Arts³, Jan Derksen³, Mike S. M. Jetten¹, Stefan Schouten⁴, Jaap S. Sinninghe Damsté⁴, Leon P. M. Lamers², Jan G. M. Roelofs², Huub J. M. Op den Camp¹ & Marc Strous¹

Wetlands are the largest natural source of atmospheric methane¹, the second most important greenhouse gas². Methane flux to the atmosphere depends strongly on the climate³; however, by far the largest part of the methane formed in wetland ecosystems is recycled and does not reach the atmosphere^{4,5}. The biogeochemical controls on the efficient oxidation of methane are still poorly understood. Here we show that submerged *Sphagnum* mosses, the dominant plants in some of these habitats, consume methane through symbiosis with partly endophytic methanotrophic bacteria, leading to highly effective *in situ* methane recycling. Molecular probes revealed the presence of the bacteria in the hyaline cells of the plant and on stem leaves. Incubation with ¹³C-methane showed rapid *in situ* oxidation by these bacteria to carbon dioxide, which was subsequently fixed by *Sphagnum*, as shown by incorporation of ¹³C-methane into plant sterols. In this way, methane acts as a significant (10–15%) carbon source for *Sphagnum*. The symbiosis explains both the efficient recycling of methane and the high organic carbon burial in these wetland ecosystems.

Peat bogs alternate between lawns and pools. Lawns are dominated by species that grow up to several decimetres above the water table. Pools are dominated by aquatic species, such as *Sphagnum cuspidatum*, that form layers of living plants below the water table. We investigated the methane-oxidizing activity of submerged *S. cuspidatum* from peat bog pools at different field locations in the Netherlands, and compared it to the activity of *S. magellanicum* and *S. papillosum* growing in lawns. The potential methane-oxidizing activity was substantially higher in the submerged mosses (Fig. 1). In control experiments with bog water, methane was not oxidized, indicating that the methanotrophic bacteria were mainly present on or in the living *Sphagnum* tissue.

The identity and location of these methanotrophs was determined in a molecular approach. Total genomic DNA from washed *Sphagnum* plants was isolated and bacterial 16S ribosomal RNA genes were amplified, cloned into *Escherichia coli*, sequenced and analysed phylogenetically. One of the 16S rRNA gene sequences of the clone library was affiliated to a cluster of type II methanotrophs that contained acidophilic methanotrophs isolated from *Sphagnum* bogs, such as *Methylocella palustris* (identity 93%)⁶ and *Methylocapsa acidiphila* (identity 93%)⁷.

The full 16S rRNA gene sequence was used to design two specific oligonucleotide probes for fluorescence *in situ* hybridization (FISH). FISH was combined with serial sectioning of the stems and the stem leaves of multiple individuals of submerged *S. cuspidatum*. The methanotrophic bacterium targeted by the probes was the dominant methanotroph in *S. cuspidatum* sections, accounting for over 75% of

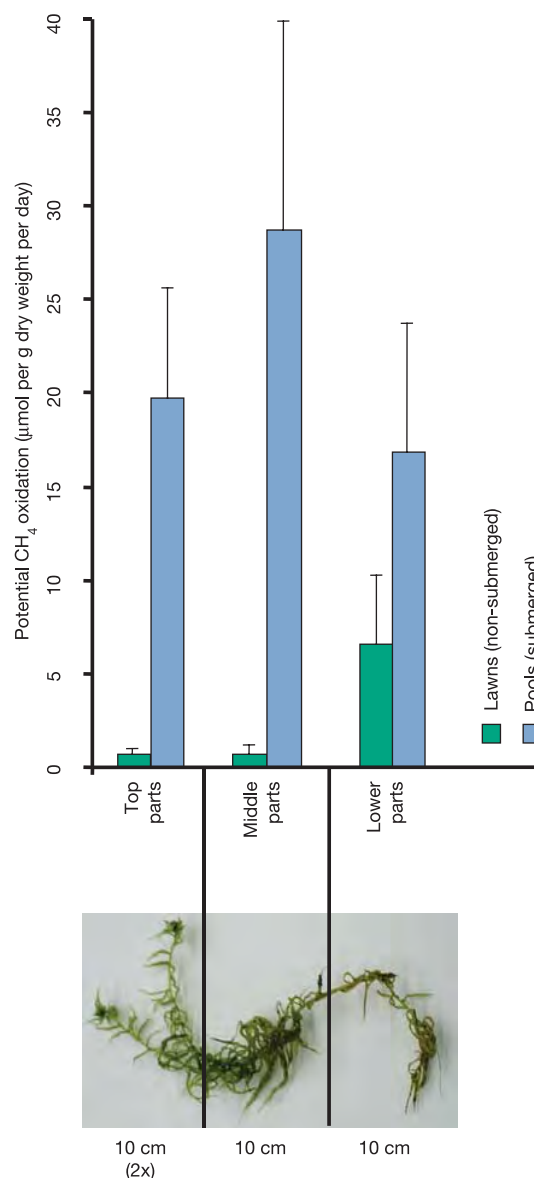


Figure 1 | Methane oxidation potential of different parts of submerged and non-submerged *Sphagnum* mosses as a measure of methanotrophs associated. Error bars indicate standard deviations of at least six independent experiments.

¹Department of Microbiology, ²Department of Aquatic Ecology and Environmental Biology, and ³Department of Plant Cell Biology, Radboud University Nijmegen, Toernooiveld 1, 6525 ED Nijmegen, The Netherlands. ⁴Royal Netherlands Institute for Sea Research (NIOZ), Department of Marine Biogeochemistry and Toxicology, PO Box 59, 1790 AB Den Burg, The Netherlands.

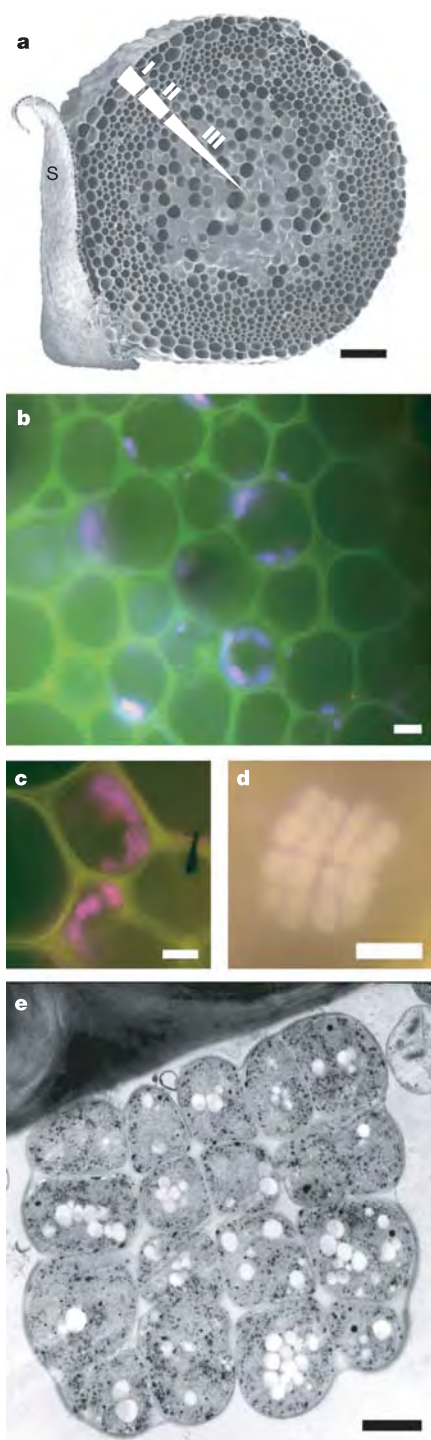


Figure 2 | *In situ* detection of the new methanotroph in *S. cuspidatum* with fluorescently labelled rRNA-targeted oligonucleotide probes. **a**, Cryo-scanning electron micrograph of a stem cross-section. S, stem leaf; I, outer cortex; II, internal cylinder; III, inner pith. Scale bar, 100 μ m. **b**, **c**, Epifluorescence micrographs of the new methanotroph (purple or pink cells) in the outer cortex of *Sphagnum* stems, after a double hybridization with the specific probe S⁺-18ALF-1437-a-A-18 and the general probe EUB²¹. Scale bars, 10 and 5 μ m. **d**, Dense, geometric clusters of the same bacterium on a stem leaf, after a triple hybridization with the specific probe S⁺-18ALF-1437-a-A-18, the general probe EUB and probe Alf968 (S-Sc-aProt-0968-a-A-18 (5'-GGTAAGGTTCTGCGCGTT-3'), specific for α -Proteobacteria). Scale bar, 5 μ m. **e**, Transmission electron micrograph of a geometric cluster closely attached to a stem leaf. Scale bar, 1 μ m.

Table 1 | Methane and CO₂ concentrations and $\delta^{13}\text{C}$ values in the Mariapeel bog pool

	CH ₄	CO ₂	Plants*
Sediment gas composition (%)	52	48	–
Bulk water concentration (μ M)	50 $\pm$ 20	160 $\pm$ 30	–
$\delta^{13}\text{C}$ (‰)	–56	–14.5	–26.5

* The $\delta^{13}\text{C}$ values of growing (–26‰) and decaying (–27‰) *S. cuspidatum* were almost identical.

all α -Proteobacteria. Application of general probes showed that the α -Proteobacteria themselves made up 80% of all detected bacteria, indicating that the new methanotroph was indeed the dominant bacterium in *S. cuspidatum* sections. γ -Proteobacteria (including type I methanotrophs) were virtually absent.

In *S. cuspidatum* stems, clusters of the new methanotroph were present in the hyaline cells of the outer cortex (Fig. 2a–c; in total 10^6 – 10^7 methanotrophs per individual plant, total length of stem $\sim$ 40 cm). Hyaline cells are dead, water-filled cells that contain pores by which solutes (and bacteria) can move in or out⁸. The presence of clusters indicated that this bacterium was actively growing inside the hyaline cells. The bacterial clusters consisted of 5–25 individual coccoid cells lying closely together in a random arrangement. On the stem leaves, the same probes hybridized with bacteria occurring as dense, geometric clusters tightly bound to the living plant cells (Fig. 2d, e, 10^5 – 10^6 methanotrophs per individual plant). Differences in the morphology of micro-colonies have been observed to depend on environmental conditions for other microorganisms⁹. On the basis of the measured *in vitro* methane-oxidizing capacity of *S. cuspidatum* ($\sim$ 20 μ mol per g dry weight per day; Fig. 1) and the number of methanotrophs per plant, an activity in the order of 1–4 fmol methane cell^{–1} h^{–1} was estimated for the associated methanotrophs. This is significantly higher than the *in vitro* methane oxidation rates reported for acidophilic methanotrophs ($\sim$ 0.3 fmol methane cell^{–1} h^{–1}) (ref. 6), indicating that the actual numbers of methanotrophs per *S. cuspidatum* individual might still be underestimated.

Because FISH analysis had shown that the new methanotroph was the only bacterium occurring in the characteristic geometric clusters, it was possible to identify and inspect this bacterium with transmission electron microscopy (TEM). The TEM and FISH results were consistent with respect to the localization of the methanotroph. TEM also showed that this bacterium did not contain any intracytoplasmic membranes. The absence of intracytoplasmic membranes was noted previously for the phylogenetically related type II methanotroph *M. palustris*⁶. Otherwise, intracytoplasmic membranes are a characteristic feature of methanotrophic bacteria.

The predominance of type II methanotrophs was further substantiated by the presence of bishomohopanoic acid in *Sphagnum* lipid extracts after periodic acid treatment. This compound was previously shown to form after periodic acid treatment from the C₃₅ hopanetetrol derivatives, membrane rigidifiers produced by methanotrophic bacteria¹⁰. The natural ^{13}C contents of this compound ($\delta^{13}\text{C}$ = –39.8‰) were substantially depleted relative to *Sphagnum* cell material and enriched compared to that of methane (Table 1), in accordance with its origin from serine-cycle (type II) methanotrophic bacteria¹¹. Using this methanotrophic biological marker we were able to determine whether the methanotrophs associated with *Sphagnum* were actively growing. After incubating *Sphagnum* with ^{13}C -labelled methane for 5 days, isotopic analysis showed that ^{13}C -labelled methane was incorporated into this lipid in substantial amounts; nearly 50% of this lipid was synthesized from the labelled methane, indicating that the methanotrophic population had doubled over the course of the experiment.

The observed tight association of methanotrophic bacteria with *S. cuspidatum* would enable the efficient recycling into living mosses

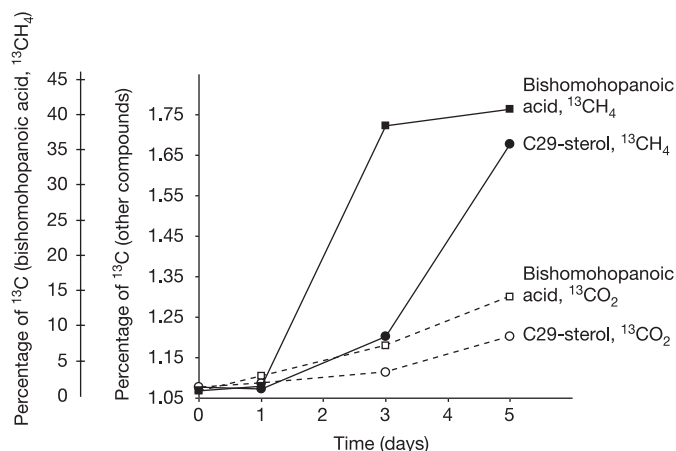
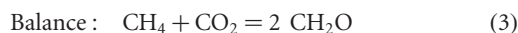
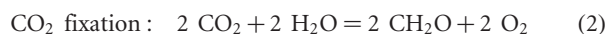


Figure 3 | Incorporation of ^{13}C label in biological markers for *Sphagnum* (circles) and methanotrophic bacteria (squares). Filled circles/solid lines show the results for labelled methane (99% ^{13}C) in the presence of unlabelled carbon dioxide; open symbols/dashed lines show the results for labelled carbon dioxide (4% ^{13}C).

of both oxygen (derived from photosynthesis) and methane (derived from decaying plants), according to the following set of equations:



To provide experimental evidence for this scenario, the potential contribution of methane to carbon fixation by *S. cuspidatum* was investigated under conditions relevant to the field. Multiple batches of individuals of *S. cuspidatum* were incubated with ^{13}C -labelled methane in the presence of unlabelled carbon dioxide. As a control experiment, only ^{13}C -labelled carbon dioxide was supplied. Both compounds were added to a final concentration of 0.2 mM, close to the *in situ* concentrations (Table 1). Over 5 days, incorporation of the label by *S. cuspidatum* was determined via the ^{13}C incorporation into sitosterol, a *Sphagnum*-specific sterol (Fig. 3). Methane was assimilated into the sitosterol pool at a rate of $0.20 \pm 0.03 \mu\text{g C per g dry weight per day}$, compared to $1.4 \pm 0.1 \mu\text{g C per g dry weight per day}$ for carbon dioxide. Thus, in the presence of carbon dioxide, at near *in situ* concentrations, the capacity of methane incorporation by *S. cuspidatum* was ~15% of the carbon dioxide assimilation capacity.

The natural carbon isotope abundances of *Sphagnum* mosses in the field ($\delta^{13}\text{C} -26.5\%$; Table 1) are consistent with our estimate that 15% of the carbon fixed by *Sphagnum* derives from isotopically depleted methane (that is, -56% ; Table 1). *S. cuspidatum* fixes carbon dioxide via the Calvin cycle, and is able to fractionate strongly against ^{13}C (up to 29‰) at high carbon dioxide concentrations ($>2 \text{ mM}$)^{12,13}. However, unlike vascular (semi-)aquatic plants such as rice, *S. cuspidatum* does not have aerenchyma⁸ that facilitate the transport of atmospheric carbon dioxide. Therefore, at lower carbon dioxide concentrations, carbon assimilation by *S. cuspidatum* is limited by mass transfer, and carbon fractionation has been reported to decrease to at most 4‰ (refs 12, 13). Because the average carbon dioxide concentration in the field was approximately 0.16 mM, a range of 4–10‰ was used as a conservative estimate for carbon fractionation by *S. cuspidatum* in the field^{12,13}. With this assumption, the data from Table 1 and a simple isotopic mass balance (see Methods), we calculated that methane contributed on average between 5% and 20% to the carbon fixed by *S. cuspidatum* in the field, in good agreement with the labelling results. It is likely that variation in local conditions (water depth, exposure to wind,

temperature, light availability, rates of methane ebullition compared to diffusion/advection) will affect the relative contribution of methane to the carbon uptake of *Sphagnum* mosses in space and time. This will also be determined by the location of the symbiotic methanotrophs in the plants, both in the direct vicinity of the photosynthetically active cells and in the more remote hyaline cells of the stems.

Our results show that methane is a significant and as yet overlooked supplement to the carbon intake of submerged *S. cuspidatum* in peat bogs. Peat bogs in the Northern Hemisphere store up to one-third of the carbon sequestered in soils globally¹⁴. This is surprising considering that the primary production is limited by the nutrient delivery through rain water and the limited delivery of carbon dioxide to the acidic waters of these ecosystems⁵. The efficient recycling of peat decomposition products (such as methane) as demonstrated here may mechanistically explain the paradox of peatlands as ecosystems with apparent low primary productivity combined with high carbon burial.

METHODS

In situ conditions. In the Maripeel nature reserve (the Netherlands: $51^\circ 24' 90'' \text{ N}$; $5^\circ 54' 90'' \text{ E}$), $\delta^{13}\text{C}$ values of *Sphagnum* mosses and material from the decaying peat were determined on freeze-dried homogenized material as described previously¹⁵. Concentrations and $\delta^{13}\text{C}$ values of dissolved carbon dioxide and methane were measured as described previously¹⁶.

Methane oxidation. Potential methane oxidation of different parts of *Sphagnum* were measured by incubating 6 g of thoroughly washed *Sphagnum* in 100 ml infusion flasks sealed with airtight rubber stoppers. To prevent mass transport limitations, no additional water was added to the experiments. To each flask 1 ml of pure methane was added and methane consumption was measured every 6 h over 2 days. Methane oxidation rates were calculated by linear regression. Tenfold concentrated water samples (10^6 bacterial cells ml^{-1}) from the bog were used as controls and showed no methane oxidation. Samples were collected in the Netherlands from seven lawn locations (*S. magellanicum*, *S. papillosum*) and six bog pools (*S. cuspidatum*), one of these being the Maripeel. Methane was measured on an HP 5890 gas chromatograph equipped with a flame ionization detector and a Porapak Q column (80/100 mesh).

16S rRNA gene sequence analysis, FISH and electron microscopy. Total genomic DNA from *S. cuspidatum* plants containing methanotrophs, isolated with combined methods¹⁷, was used as template for PCR amplification of 16S rRNA genes. PCR was performed with general bacterial primers¹⁸ using a T gradient thermal cycler (Biometra), and a clone library was made as described previously¹⁸. Based on the obtained 16S rRNA gene sequences, two new oligonucleotide probes S*-18ALF-0218-a-A-18 (5'-GGGCCGATCCCCCGGCGA-3') and S*-18ALF-1437-a-A-18 (5'-CTTGCGGTAAACAGACG-3') were designed using the ARB program package¹⁹. Apart from these species-specific probes we used group-specific probes described previously^{20,21}. Fresh *S. cuspidatum* stems sectioned with a scalpel (section thickness $0.1 \pm 0.05 \text{ mm}$) were used for FISH as described previously²². Formaldehyde concentrations used in the FISH experiments varied between 10% and 20%. No signal was obtained at these formaldehyde concentrations when testing the specificity of the probes with *Beijerinckia indica* ssp. *indica* (DSM 1715), which has the fewest mismatches of all reference organisms to the designed probes. Electron microscopy (TEM/SEM) was performed on stems and stem leaves following published protocols^{6,22}.

Methane incorporation measurements. *S. cuspidatum*, collected from the Maripeel nature reserve, was washed with demineralized water and incubated in 250-ml serum bottles in 5 g wet weight aliquots with 150 ml medium as described previously⁸. ^{13}C - or ^{12}C - CH_4 or CO_2 were added to final concentrations of 200 μM as specified in the text. The bottles were shaken at 150 r.p.m. at ambient conditions and sacrificed for lipid analysis at days 0, 1, 3 and 5. Lipids were ultrasonically extracted and analysed by gas chromatography/mass spectrometry and isotope ratio gas chromatography mass spectrometry as described²³. Hopanes were analysed by treatment of the lipid fraction with periodic acid and sodium borohydride as described previously^{10,24}.

Isotopic mass balancing. The measured $\delta^{13}\text{C}$ values of *S. cuspidatum* in the field (-26% , see Table 1) resulted from assimilation of dissolved carbon dioxide (-14.5%), respired methane (-56%) and fractionation against ^{13}C during (mass-transfer-limited) carbon dioxide fixation^{12,13} ($7 \pm 3\%$). The following equation describes this relationship quantitatively (where *a* denotes the fraction of plant carbon derived from methane and *Ep* denotes the fractionation during

fixation):

$$\delta^{13}\text{C}(\text{Sphagnum}) = a\delta^{13}\text{C}(\text{respired methane}) + (1 - a)\delta^{13}\text{C}(\text{carbon dioxide}) - E_p \quad (4)$$

Because all $\delta^{13}\text{C}$ values from equation (4) were known experimentally, it could be derived that the contribution of methane to *Sphagnum* carbon (a) was between 0.05 and 0.2 (equivalent to 5–20%).

Received 23 December 2004; accepted 9 May 2005.

- Hein, R., Crutzen, P. J. & Heimann, M. An inverse modeling approach to investigate the global atmospheric methane cycle. *Global Biogeochem. Cycles* **11**, 43–76 (1997).
- Rodhe, H. A comparison of the contribution of various gases to the greenhouse effect. *Science* **248**, 1217–1219 (1990).
- Smith, L. C. *et al.* Siberian peatlands a net carbon sink and global methane source since the early Holocene. *Science* **303**, 353–356 (2004).
- Dedysh, S. N. *et al.* Isolation of acidophilic methane-oxidizing bacteria from northern peat wetlands. *Science* **282**, 281–284 (1998).
- Lamers, L. P. M., Farhoush, C., Van Groenendael, J. M. & Roelofs, J. G. M. Calcareous groundwater raises bogs; the concept of ombrotrophy revisited. *J. Ecol.* **87**, 639–648 (1999).
- Dedysh, S. N. *et al.* *Methylocella palustris* gen. nov., sp. nov., a new methane-oxidizing acidophilic bacterium from peat bogs, representing a novel subtype of serine-pathway methanotrophs. *Int. J. Syst. Evol. Microbiol.* **50**, 955–969 (2000).
- Dedysh, S. N. *et al.* *Methylocapsa acidophila* gen. nov., sp. nov., a novel methane-oxidizing and dinitrogen fixing acidophilic bacterium from *Sphagnum* bog. *Int. J. Syst. Evol. Microbiol.* **52**, 251–261 (2002).
- Rydin, H. & Clymo, R. S. Transport of carbon and phosphorus compounds about *Sphagnum*. *Proc. R. Soc. Lond.* **237**, 63–84 (1989).
- Yao, R., Macario, A. J. L. & Conway de Macario, E. Immunochemical differences among *Methanosarcina mazei* S-6 morphologic forms. *J. Bacteriol.* **174**, 4683–4688 (1992).
- Rohmer, M., Bissleret, P. & Neunlist, S. in *Biological Markers in Sediments and Petroleum* (eds Moldowan, J. M., Albrecht, P. & Philp, R. P.) 1–17 (Prentice Hall, London, 1992).
- Jahnke, L. L., Summons, R. E., Hope, J. M. & Des Marais, D. J. Carbon isotopic fractionation in lipids from methanotrophic bacteria II: The effects of physiology and environmental parameters on the biosynthesis and isotopic signatures of biomarkers. *Geochim. Cosmochim. Acta* **63**, 79–93 (1999).
- Keeley, J. E. & Sandquist, D. R. Carbon: freshwater plants. *Plant Cell Environ.* **15**, 1021–1035 (1992).
- Smolders, A. J. P., Tomassen, H. B. M., van Mullekom, M., Lamers, L. P. M. & Roelofs, J. G. M. Mechanisms involved in the re-establishment of *Sphagnum*-dominated vegetation in rewetted bog remnants. *Wetlands Ecol. Manag.* **11**, 403–418 (2003).
- Post, W. M., Emanuel, W. R., Zinke, P. J. & Strangenberger, A. G. Soil carbon pools and world life zones. *Nature* **298**, 156 (1982).
- Marguillier, S., van der Velde, G., Dehairs, F., Hemminga, M. A. & Rajagopal, S. Trophic relationship in an interlinked mangrove-seagrass ecosystem as traced by $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$. *Mar. Ecol. Prog. Ser.* **151**, 115–121 (1997).
- Smolders, A. J. P., Tomassen, H. B. M., Lamers, L. P. M., Lomans, B. P. & Roelofs, J. G. M. Peat bog restoration by floating raft formation: the effects of groundwater and peat quality. *J. Appl. Ecol.* **39**, 391–401 (2002).
- Lomans, B. P. *et al.* Microbial populations involved in cycling of dimethyl sulfide and methanethiol in freshwater sediments. *Appl. Environ. Microbiol.* **67**, 1044–1051 (2001).
- Juretschko, S. *et al.* Combined molecular and conventional analyses of nitrifying bacterium diversity in activated sludge: *Nitrosococcus mobilis* and *Nitrospira*-like bacteria as dominant populations. *Appl. Environ. Microbiol.* **64**, 3042–3051 (1998).
- Ludwig, W. *et al.* ARB: A software environment for sequence data. *Nucleic Acids Res.* **32**, 1363–1371 (2004).
- Amann, R. I. *et al.* Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl. Environ. Microbiol.* **56**, 1919–1925 (1990).
- Daims, H., Bruhl, A., Amann, R., Schleifer, K. H. & Wagner, M. The domain-specific probe EUB338 is insufficient for the detection of all Bacteria: development and evaluation of a more comprehensive probe set. *Syst. Appl. Microbiol.* **22**, 434–444 (1999).
- Wolters-Arts, M. *et al.* Water-conducting properties of lipids during pollen hydration. *Plant Cell Environ.* **25**, 513–519 (2002).
- Schouten, S. *et al.* Biosynthetic effects on the stable carbon isotopic compositions of algal lipids: Implications for deciphering the carbon isotopic biomarker record. *Geochim. Cosmochim. Acta* **62**, 1397–1406 (1998).
- Sinninghe Damsté, J. S. *et al.* The occurrence of hopanoids in planctomycetes: implications for the sedimentary biomarker record. *Organic Geochim.* **35**, 561–566 (2004).

Acknowledgements We thank K. T. van de Pas-Schoonen, A. Pol, H. P. M. Geurts, J. Eygensteyn, M. van Mullekom, J. Berk, H. Tomassen and M. M. A. van Herpen for technical support. Part of this study was supported by the Dutch Ministry of Agriculture, Nature Management and Food quality (Research Program 'Overlevingsplan Bos en Natuur').

Author Information The 16S rRNA gene sequences were deposited at GenBank under accession number AY163571. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.J.P.S. (a.smolders@science.ru.nl) or J.S.S.D. (damste@nioz.nl).

The contribution of species richness and composition to bacterial services

Thomas Bell^{1,2}, Jonathan A. Newman³, Bernard W. Silverman⁴, Sarah L. Turner² & Andrew K. Lilley²

Bacterial communities provide important services. They break down pollutants, municipal waste and ingested food, and they are the primary means by which organic matter is recycled to plants and other autotrophs. However, the processes that determine the rate at which these services are supplied are only starting to be identified. Biodiversity influences the way in which ecosystems function¹, but the form of the relationship between bacterial biodiversity and functioning remains poorly understood. Here we describe a manipulative experiment that measured how biodiversity affects the functioning of communities containing up to 72 bacterial species constructed from a collection of naturally occurring culturable bacteria. The experimental design allowed us to manipulate large numbers of bacterial species selected at random from those that were culturable. We demonstrate that there is a decelerating relationship between community respiration and increasing bacterial diversity. We also show that both synergistic interactions among bacterial species and the composition of the bacterial community are important in determining the level of ecosystem functioning.

Two principal mechanisms underlie our current understanding of how biodiversity affects ecosystem functioning, and especially those functions related to ecosystem productivity (the rate of biomass turnover)². First, different species use slightly different resources. Species-rich communities are therefore more productive because more of the overall resource is used (the 'complementarity mechanism')². Second, there is variation in the magnitude of individual species' effects on ecosystem functioning. Species-rich communities are therefore more productive on average because they are more likely to contain species with a large effect on ecosystem functioning (the 'selection mechanism')². Both mechanisms predict a decelerating diversity–functioning relationship under some conditions (Fig. 1), so it is not possible to distinguish between the two mechanisms on the basis of the shape of the diversity–function relationship alone. Both of these mechanisms are important in determining the level of ecosystem functioning^{3,4}, but their relative importance seems to depend on the particular ecosystem under investigation. Complementarity is thought to be relatively unimportant in natural bacterial communities^{5–7}. Although it is possible to manufacture communities in which the constituent species are complementary, many bacterial species are functionally redundant when tested on individual substrates⁸. The addition of species to species-poor communities is therefore unlikely to have a substantial effect on the level of ecosystem functioning, except by increasing the probability of selecting species that contribute greatly to functioning^{9–11}.

Experiments to test these ideas have been hampered by methodological difficulties associated with manipulating complex natural bacterial communities. Laboratory microcosm experiments, which

construct communities from pure cultures, have manipulated only a few (<20) readily available species¹². Such experiments indicate which mechanisms are possible, but not which mechanisms are important under natural conditions. More diverse intact bacterial communities can be manipulated by applying varying levels of a stress (for example, a bactericide)¹³, by diluting and re-growing a community^{5,7}, or by making use of natural differences in community composition^{10,11,14}. Unfortunately, in such experiments the species composition of the low diversity community is a subset of the high diversity community, so species composition (types of species) is confounded with species richness (number of species)¹⁵. Here we report the results of an experiment that manipulated relatively large numbers of culturable bacterial species selected at random from a naturally occurring aerobic bacterial assemblage using a new experimental design that enables us to separate the contributions of species richness and species composition to bacterial services.

The ecosystems that we use are the semi-permanent rainpools that form in bark-lined depressions near the base of large European beech trees (*Fagus sylvatica*). These natural microcosms house an array of heterotrophic organisms, the energy for which is derived principally from beech leaf litter¹⁶. Bacterial diversity and composition vary substantially among treeholes¹⁷, so it is reasonable to ask whether this natural variation results in differences in ecosystem functioning. We isolated bacteria from six treeholes by randomly picking colonies that grew on nutrient agar, and determined whether the isolates were of the same or different species using standard identification protocols.

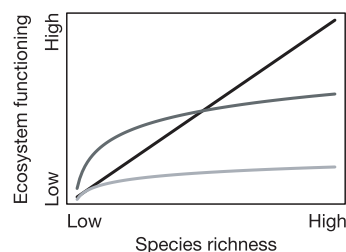


Figure 1 | The relationship between species richness and ecosystem functioning. Illustration of how the complementarity mechanism and the selection mechanism alter the shape of the relationship between species richness and some measure of ecosystem functioning. If all species contribute approximately equally to ecosystem functioning, species effects are (1) additive if the species are completely complementary (black line) or (2) decelerating if the species are to some extent functionally redundant (light grey line). If the same pool of species now contains a few species that, when present in a given mixture, are able to attain maximum ecosystem functioning, the shape of the curve will also be decelerating (dark grey line).

¹Department of Zoology, University of Oxford, South Parks Road, Oxford, OX1 3PS, UK. ²Molecular Microbial Ecology Group, NERC Centre for Ecology and Hydrology, Mansfield Road, Oxford, OX1 3SR, UK. ³Department of Environmental Biology, University of Guelph, Guelph, Ontario N1G 2W1, Canada. ⁴St. Peter's College, Oxford, OX1 2DL, UK.

The 347 isolates were composed of 103 distinct species level identities, of which we selected 72 at random to use in the current experiment. This procedure selected only for species that are aerobic heterotrophs capable of forming colonies on agar plates. As such, the results cannot reliably be generalized to the larger treehole bacterial community, except insofar as they demonstrate what results are possible for the larger community.

Microcosms consisting of sterile beech leaf disks and 10 ml of liquid (phosphate buffer) were inoculated with random combinations of the 72 selected bacterial species. Ecosystem functioning was measured as the daily respiration rate of the bacterial community in each of 1,374 microcosms over three time intervals (days 0–7, 7–14 and 14–28). We chose respiration as our measure of ecosystem functioning for two reasons. First, much of the work on ecosystem functioning in other systems has concentrated on biomass turnover, which is closely related to respiration. Second, this method automatically integrates respiration over several days, and so is not affected by transient dynamics, as would be the case for example if we measured standing biomass at the conclusion of the experiment.

The mean daily rate of bacterial respiration increased rapidly with increasing species richness from monocultures ($5.07 \pm 0.22 \mu\text{g CO}_2 \text{ ml}^{-1} \text{ d}^{-1}$, mean $\pm$ s.e.m. throughout) to 18 species ($8.93 \pm 0.42 \mu\text{g CO}_2 \text{ ml}^{-1} \text{ d}^{-1}$), but the increase slowed as species richness increased to 72 species ($11.35 \pm 0.96 \mu\text{g CO}_2 \text{ ml}^{-1} \text{ d}^{-1}$) (Fig. 2). This decelerating curve is well described by a linear dependence of respiration rate on log-transformed species richness, and as such is similar to the relationship found for other groups of organisms^{3,18–21}. The data from the three time periods are well described by similar relationships, but the absolute amount of respiration at each level of diversity and the slope of the curves declined over time (Fig. 3). Notably, the shape of the curve suggests that ecosystem functioning does not reach an asymptote until well beyond 72 species, indicating that further increases in diversity will continue to increase the level of ecosystem functioning.

We evaluated the effect of species composition and species richness on respiration using a series of linear models in which the effects of time (T), species richness (R), species composition (C), and their interactions were entered sequentially (Table 1). In these models, species richness reflects the average contribution of the number of species to respiration, irrespective of the particular species that are present. Similarly, species composition reflects the average contribution of each species to respiration, irrespective of the level of species richness, summed over every species. The model coefficients associated with the presence of each species give an estimate of the effect of the species on respiration. Each species appeared with equal

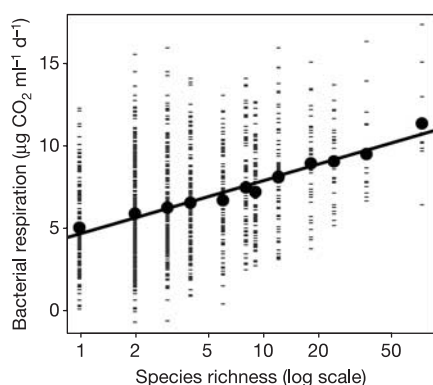


Figure 2 | Relationship between manipulated species richness (R) and ecosystem functioning (F) over 28 days. F represents the mean daily rate of bacterial community respiration. Each data point (denoted by a dash) is the mean of two replicates with identical species composition ($n = 687$). Filled circles are the means at each level of species richness ($F = 1.21 \ln(R) + 4.98$; $R^2 = 0.12$).

frequency at each level of species richness, which excluded the possibility that any increase in respiration with increasing species richness was due to the chance selection at high levels of species richness of species with a large effect on respiration.

Inspection of the sums of squares indicates that, as implied by Fig. 3, much of the variability in bacterial respiration rates is due to time, presumably because only recalcitrant resource remained after the first week. The linear effect of species richness represents the situation in which every species is identical and has additive effects. The nonlinear effect of species richness (that is, $\log(R)$, the log-transformed species richness, see step 3 in Table 1) represents the effect of richness over and above the additive effects. $\log(R)$ therefore describes how the effect of richness deviates from complete complementarity (Fig. 1) and consequently is the effect that is of most interest in studies such as this. Each species is equally represented at each level of species richness, and therefore $\log(R)$ describes the effect of interactions among species on bacterial respiration. This effect is clearly significant and the slope changes with time (see Fig. 3 and steps 3, 4 in Table 1). Inspection of Fig. 2 indicates that some of the species-poor microcosms performed as well or better than the most species-rich microcosms, suggesting that species composition is also important in determining the overall rate of respiration. This is confirmed by our analysis, showing that species composition had a smaller but significant effect on the respiration rate that interacts with species richness (steps 5, 6 in Table 1). Unlike similar previous studies, in this experiment the effects of species composition and log-transformed species richness are (at this point) orthogonal, and therefore the order of entry into the linear model does not affect the conclusions.

The coefficient associated with each of the species indicates their contribution to ecosystem functioning. The coefficients are approximately normally distributed (Fig. 4), showing that most species contributed only marginally and that no single species dominated the respiration rate. Detailed consideration of the species' effects and their interaction with diversity (see Supplementary Discussion) indicates that there is a fairly weak but significant correlation ($r = -0.42$) between the species effects (C) and the $C \times \log(R)$ interaction. On the whole, species that have high effects at moderate diversity tend to have higher effects at lower diversity and lower effects at higher diversity (Supplementary Fig. S1). However, this is only a general trend and cannot be deduced for all species individually.

Synergistic interactions among bacterial species, of which complementarity is one possibility, had an important role in determining the rate of functioning in this ecosystem, and therefore

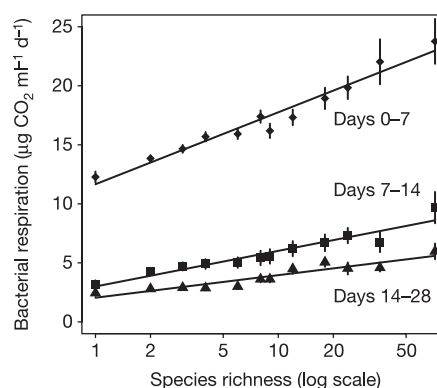


Figure 3 | Relationship between manipulated species richness (R) and ecosystem functioning (F) over each of the three time periods. F is the mean daily rate of bacterial respiration $\pm$ s.e.m. Each data point is first averaged over two replicates identical in their composition, and then across all data at each level of species richness. Error bars are standard errors (days 0–7; $F = 2.32 \ln(R) + 12.18$, $R^2 = 0.13$; days 7–14; $F = 1.12 \ln(R) + 3.33$, $R^2 = 0.03$; days 14–28; $F = 0.7 \ln(R) + 2.20$, $R^2 = 0.05$).

Table 1 | Linear models estimating the effect of time, species richness and species composition on bacterial respiration

Step	Model	Res. d.f.	Res. SS	Treat. d.f.	Treat. SS	F	AIC
1	Intercept	4,121	283,519	—	—	—	27,236.9
2	1 + T + R + T × R	4,116	115,026	5	168,493	1,269.4*	23,528.4
3	2 + log(R)	4,115	113,528	1	1,499	56.5*	23,476.3
4	3 + T × log(R)	4,113	112,916	2	612	11.5*	23,458.1
5	4 + C	4,042	108,756	71	4,160	2.2*	23,445.3
6	5 + C × log(R)	3,970	104,732	72	4,024	2.1*	23,433.9
7	6 + T × C	3,828	101,622	142	3,110	0.8	23,593.7

The linear models describe the effect of time (T), species richness (R) and species composition (C) on the daily rate of microbial respiration ($\mu\text{g CO}_2 \text{ ml}^{-1} \text{ d}^{-1}$). At each step (1–7), terms are added to the linear model and the residual degrees of freedom (Res. d.f.) and sum of squares (Res. SS) are re-calculated. The treatment degrees of freedom (Treat. d.f.), sum of squares (Treat. SS), and *F*-statistic (*F*) are calculated at each step only for the term that has been added to the model during that step. Asterisk denotes $P < 0.0001$. Akaike's 'an information criterion' (AIC) is calculated for each model. Lower AIC values indicate an improved model.

contrast with previous studies that have emphasized the role of community composition over the role of diversity in determining bacterial community functioning^{6,10,11,13,14}. The selection mechanism would be relatively unimportant in this system because no single species dominated the respiration rate (Fig. 4). Nonetheless, differences in the species composition effects were important in determining the level of respiration, and as such, the results are in general agreement with studies that have estimated the effect of community composition using observed differences in natural communities^{10,11,14}. However, composition seems to play a more minor role relative to species richness (Table 1). The data demonstrate that it is possible in principle to engineer a bacterial community to obtain the maximum rate of ecosystem functioning by selecting a particular consortium of species, as is already the practice for some industrial applications²². Enhancing the biodiversity of these systems might provide a more rapid and equally reliable solution. Positive interactions can also alter the shape of the diversity–function relationship²³, and might be particularly important in bacterial communities in which resource is often processed by a number of different species. We do not discuss this possibility in detail because the experiment was not designed to detect such interactions, but the mechanism remains an interesting prospect for future research.

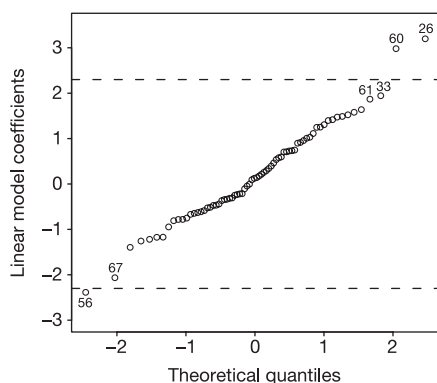


Figure 4 | Linear model coefficients as a function of the theoretical quantiles of the normal distribution. Each data point is the linear model coefficient associated with the 72 species used in the experiment. The numbers over the most significant data points refer to the species codes used in Supplementary Table S1. Positive coefficients indicate that the species has an above-average contribution to respiration, negative values denote a below-average contribution. A linear relationship indicates that the coefficients are normally distributed. Individual *t*-values test whether each of the coefficients differs from zero. The standard error of each of the species' coefficients is approximately equal (0.68 ± 0.001 , mean $\pm$ s.e.m.), and so the significance thresholds for each species are also equal. The data points between the dashed lines are not significantly different from zero following a Bonferroni correlation for multiple tests ($\alpha = 0.05/72 = 6.9 \times 10^{-4}$). The coefficient for the log-transformed species richness factor in the same linear model is 1.77 ($P < 10^{-5}$).

Our results cannot be directly generalized to natural environments. The bacterial species that are apparently unimportant in relatively stable microcosm environments might have an important role in maintaining natural levels of community respiration when conditions fluctuate. Here we have chosen only aerobic heterotrophs that are culturable on R2 agar, so large portions of the treehole bacterial community are not considered, including for example the obligate anaerobes. Even with extensive culturing on multiple media, typically only a fraction ($< 10\%$) of bacterial species are culturable²⁴. Several studies have now compared the culturable to the unculturable community^{24,25}, and there seems to be a consensus that they often bear little resemblance to each other. Our library of bacterial strains is likely to be a similarly unrepresentative sample of those discovered using molecular techniques, in the sense that the species that dominate clone libraries will be different from those that dominate our library of culturable bacteria. We would therefore caution against extending our results to treehole bacterial communities in general. However, it is clear from the FAME identifications (see Methods and Supplementary Table S1) that we have obtained a broad cross-section of bacterial taxonomic and functional richness. Although it is not possible to predict directly the physiological traits of our experimental strains from their nearest taxonomic neighbours (Supplementary Table S1), it is apparent that the 72 species of aerobic heterotrophic treehole bacteria chosen for this study represent a broad taxonomic range, with 27 species in 17 Gram-positive genera and 19 species in 14 Gram-negative genera. The experiment therefore provides suggestive evidence that similar processes may operate in natural environments.

The level of complexity of bacterial communities, with perhaps thousands of bacterial species contained within a few millilitres of pond water, has prevented a comprehensive manipulation of bacterial assemblages. Molecular techniques have recently enabled relatively detailed descriptions of bacterial community composition and diversity, but have not been accompanied by parallel methods to manipulate diversity. These unculturable bacteria clearly have an important role in bacterial community dynamics, but experiments remain impossible as long as the technology to manipulate individual species does not exist. Environmental microbiology can identify the potential effects of differences in composition and diversity on ecosystem functioning²⁶, but we believe that a mechanistic understanding of microbial communities will require similar large-scale manipulations of bacterial assemblages.

METHODS

Laboratory methods. We spread $20 \mu\text{l}$ serial dilutions from 6 stirred treeholes onto replicate 145 mm R2 agar (Oxoid Ltd.) plates. After incubation for 7 to 10 days at around 20°C , the resulting colonies were picked at random into 0.7 ml $10\% \text{ v/v}$ nutrient broth (Oxoid Ltd.) These isolates were assigned identities from their cellular fatty acid profiles using standard fatty acid methyl ester (FAME) extraction protocols²⁷, gas chromatography and the Sherlock Microbial Identification System (MIDI Inc.)²⁸. The 347 isolates were composed of 103 species (euclidean distance > 10), from which we selected 72 species at random for use in the current experiment (Supplementary Table S1). We designate our operational taxonomic units as 'species' as a convenient label, not because we are confident

that they are analogous to species as defined for larger organisms. We refer the reader to reviews of the bacterial species concept²⁹ and FAME identification procedures for more complete discussions of the subject²⁸.

Bacterial communities were assembled in 30-ml screw-top vials. Each of these microcosms contained 10 ml of sterile PBS buffer (pH 7.0) and 50 sterile beech leaf discs (diameter, 8 mm). The beech leaves were obtained from freshly fallen leaves collected on 18 October 2003 and stored at 4 °C until the start of the experiment. Each microcosm was inoculated with a total of 100 µl of pure culture that had been grown in 10% nutrient broth for at least one week at 23 °C (for example, four-species combinations were inoculated with 25 µl of pure culture from each of the constituent species). Although stationary phase density varied among the isolates, it was assumed that this would have little effect on the dynamics over 28 days. Approximately 50 microcosms were inoculated per day on weekdays from 26 July to 3 September 2004. The particular combinations were inoculated in a random order to prevent possible effects due to the day of the week. Each microcosm was kept at 23 °C throughout the experiment and was dismantled 28 days after the day on which it was inoculated, which is a comparable length of time to similar microcosm experiments (for example, ref. 12).

We used bacterial respiration as a measure of ecosystem functioning. Sterile 5-ml vials containing 2 ml 0.2 M NaOH solution were directly inserted into each microcosm as it was inoculated. We removed the 2 ml of NaOH after 7, 14 and 28 days, and replaced the removed NaOH with freshly prepared 0.2 M NaOH. The CO₂ respired by the bacterial communities, as well as the CO₂ sealed in the headspace of the vials, reacts with the NaOH to form Na₂CO₃ and H₂O. Titrating the NaOH against 0.02 M HCl to neutral and subtracting sterile negative controls gives the total amount of CO₂ respired by the bacterial community over the given time period³⁰.

Experimental design and statistical analysis. We assembled random combinations of species at 12 levels of species richness, because it is logistically impractical to use all possible combinations of 72 species ($>4.7 \times 10^{21}$ combinations). We used an experimental design that enabled separation of the effects of species richness and species composition. The basic building block of the experimental design is a set of 72/s microcosms, each with *s* species present. Within this set, the microcosms are constructed by sampling the 72 species without replacement. For example, if *s* = 4, we randomly partition the 72 species into 18 4-species combinations. The process of constructing a system of microcosms is carried out independently five times, and for each set of microcosms thus constructed, two independent replicates are carried out. Therefore, for any given *s*, the number of microcosms considered is $5 \times 2 \times (72/s)$. We chose the level *s* to be every factor of 72 (*s* = 1, 2, 3, 4, 6, 8, 9, 12, 18, 24, 36 and 72). A summary of the design is given in Supplementary Table S2.

The time (three levels), species richness (linear and log-transformed) and species composition (72 factors each with two levels) were entered into a series of linear models. We calculated the significance of the factors added to each model and compared the models. The third time period (days 14–28) was weighted by a factor of 4 to compensate for unequal variances among the levels of time. The model fit was improved by using diversity as a log-transformed continuous variable rather than a categorical variable.

Received 22 April; accepted 7 June 2005.

1. Loreau, M. *et al.* Biodiversity and ecosystem functioning: Current knowledge and future challenges. *Science* **294**, 804–808 (2001).
2. Loreau, M. & Hector, A. Partitioning selection and complementarity in biodiversity experiments. *Nature* **412**, 72–76 (2001).
3. Emmerson, M. C., Solan, M., Emes, C., Paterson, D. M. & Raffaelli, D. Consistent patterns and the idiosyncratic effects of biodiversity in marine ecosystems. *Nature* **411**, 73–77 (2001).
4. Hooper, D. U. *et al.* Effects of biodiversity on ecosystem functioning: A consensus of current knowledge. *Ecol. Monogr.* **75**, 3–35 (2005).
5. Garland, J. L. & Lehman, R. M. Dilution/extinction of community phenotypic characters to estimate relative structural diversity in mixed communities. *FEMS Microbiol. Ecol.* **30**, 333–343 (1999).
6. Griffiths, B. S. *et al.* An examination of the biodiversity–ecosystem function relationship in arable soil microbial communities. *Soil Biol. Biochem.* **33**, 1713–1722 (2001).
7. Franklin, R. B., Garland, J. L., Bolster, C. H. & Mills, A. L. Impact of dilution on microbial community structure and functional potential: Comparison of numerical simulations and batch culture experiments. *Appl. Environ. Microbiol.* **67**, 702–712 (2001).

8. Yin, B., Crowley, D., Sparovek, G., De Melo, W. J. & Borneman, J. Bacterial functional redundancy along a soil reclamation gradient. *Appl. Environ. Microbiol.* **66**, 4361–4365 (2000).
9. Fierer, N., Schimel, J. P. & Holden, P. A. Influence of drying–rewetting frequency on soil bacterial community structure. *Microb. Ecol.* **45**, 63–71 (2003).
10. Cavigelli, M. A. & Robertson, G. P. The functional significance of denitrifier community composition in a terrestrial ecosystem. *Ecology* **81**, 1402–1414 (2000).
11. Cavigelli, M. A. & Robertson, G. P. Role of denitrifier diversity in rates of nitrous oxide consumption in a terrestrial ecosystem. *Soil. Biol. Biochem.* **33**, 297–310 (2001).
12. Naeem, S., Hahn, D. R. & Schuurman, G. Producer–decomposer co-dependency influences biodiversity effects. *Nature* **403**, 762–764 (2000).
13. Griffiths, B. S. *et al.* The relationship between microbial community structure and functional stability, tested experimentally in an upland pasture soil. *Microb. Ecol.* **47**, 104–113 (2004).
14. Horz, H. P., Barbrook, A., Field, C. B. & Bohannon, B. J. M. Ammonia-oxidizing bacteria respond to multifactorial global change. *Proc. Natl. Acad. Sci. USA* **101**, 15136–15141 (2004).
15. Huston, M. A. Hidden treatments in ecological experiments: Re-evaluating the ecosystem function of biodiversity. *Oecologia* **110**, 449–460 (1997).
16. Kitching, R. L. An ecological study of water-filled tree-holes and their position in the woodland ecosystem. *J. Anim. Ecol.* **40**, 281–302 (1971).
17. Bell, T. *et al.* Larger islands house more bacterial taxa. *Science* **308**, 1884 (2005).
18. van der Heijden, M. G. A. *et al.* Mycorrhizal fungal diversity determines plant biodiversity, ecosystem variability and productivity. *Nature* **396**, 69–72 (1998).
19. McGrady-Steed, J., Harris, P. M. & Morin, P. J. Biodiversity regulates ecosystem predictability. *Nature* **390**, 162–165 (1997).
20. Tilman, D. *et al.* Diversity and productivity in a long-term grassland experiment. *Science* **294**, 843–845 (2001).
21. Hector, A. *et al.* Plant diversity and productivity experiments in European grasslands. *Science* **286**, 1123–1127 (1999).
22. Curtis, T. F., Head, I. M. & Graham, D. Theoretical ecology for engineering biology. *Environ. Sci. Technol.* **37**, 64A–70A (2003).
23. Cardinale, B. J., Palmer, M. A. & Collins, S. L. Species diversity enhances ecosystem functioning through interspecific facilitation. *Nature* **415**, 426–429 (2002).
24. Zengler, K. *et al.* Cultivating the uncultured. *Proc. Natl. Acad. Sci. USA* **99**, 15681–15686 (2002).
25. Rappe, M. S. & Giovannoni, S. J. The uncultured microbial majority. *Annu. Rev. Microbiol.* **57**, 369–394 (2003).
26. Schimel, J. P. & Gulledege, J. Microbial community structure and global trace gases. *Global Change Biol.* **4**, 745–758 (1998).
27. Sasser, M. MIDI Inc. technical note 101: Identification of bacteria by gas chromatography of cellular fatty acids (Newark, 2001).
28. Thompson, I. P., Bailey, M. J., Ellis, R. J. & Purdy, K. J. Subgrouping of bacterial populations by cellular fatty-acid composition. *FEMS Microbiol. Ecol.* **102**, 75–84 (1993).
29. Cohan, F. M. What are bacterial species? *Annu. Rev. Microbiol.* **56**, 457–487 (2002).
30. Zibilske, L. M. *Methods of Soil Analysis, Part 2. Microbiological and Biochemical Properties* 835–863 (SSSA, Madison, 1994).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to J. Fryxell, T. Nudds and their graduate students for providing comments on the original manuscript, to A. Singer for help developing the technique to measure bacterial respiration, and the Centre for Ecology and Hydrology in Oxford for providing the laboratory space. T. B. was supported by Fonds quebecois de la recherche sur la nature et les technologies, the Natural Sciences and Engineering Council of Canada, and the Clarendon Fund (Oxford University).

Author contributions The experiment was originally conceived by T.B., J.A.N. and A.K.L. The laboratory work was conducted by T.B. with the help of A.K.L. and S.L.T. The experimental design was conceived by A.K.L. and developed by T.B., J.A.N. and B.W.S. The statistical analyses were performed by B.W.S. and T.B. The manuscript was written principally by T.B. with extensive input from J.A.N., B.W.S. and A.K.L.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.K.L. (akl@ceh.ac.uk).

The neuronal representation of pitch in primate auditory cortex

Daniel Bendor¹ & Xiaoqin Wang¹

Pitch perception is critical for identifying and segregating auditory objects¹, especially in the context of music and speech. The perception of pitch is not unique to humans and has been experimentally demonstrated in several animal species^{2,3}. Pitch is the subjective attribute of a sound's fundamental frequency (f_0) that is determined by both the temporal regularity and average repetition rate of its acoustic waveform. Spectrally dissimilar sounds can have the same pitch if they share a common f_0 . Even when the acoustic energy at f_0 is removed ('missing fundamental') the same pitch is still perceived¹. Despite its importance for hearing, how pitch is represented in the cerebral cortex is unknown. Here we show the existence of neurons in the auditory cortex of marmoset monkeys that respond to both pure tones and missing fundamental harmonic complex sounds with the same f_0 , providing a neural correlate for pitch constancy¹. These pitch-selective neurons are located in a restricted low-frequency cortical region near the anterolateral border of the primary auditory cortex, and is consistent with the location of a pitch-selective area identified in recent imaging studies in humans^{4,5}.

Many natural sounds (or biologically significant sounds) have periodic acoustical waveforms. These sounds can be spectrally decomposed into a sinusoid at the frequency of periodicity (f_0) and a series of sinusoids at frequencies that are integer multiples of f_0 (harmonics). Although these individual spectral components are represented within the cochleotopic organization of the auditory system in a distributed fashion, they are perceptually grouped together into a single sound with a pitch equivalent to a pure tone at f_0 (ref. 1). In the auditory periphery, the f_0 of complex sounds—such as missing fundamental harmonic complex sounds (MFs)—is represented by a distributed neural code involving both the discharge rates and temporal firing patterns of auditory nerve fibres^{6,7}. How this information is used to encode pitch within the central auditory system is poorly understood.

Deficits in pitch discrimination have been observed in animals⁸, including humans^{9,10}, following auditory cortical lesions, indicating a cortical role in pitch perception. However, electrophysiological recordings in macaque monkeys suggest that primary auditory cortex (AI) does not contain a representation of pitch, as AI neurons do not respond to MFs with a pitch matching their characteristic frequency^{11,12}. Alternatively, pitch may be processed in non-primary auditory cortex, as recent human imaging studies have revealed a cortical pitch processing region anterolateral to primary auditory cortex^{4,5}. The organization of primary and secondary areas of auditory cortex is largely conserved across primate species^{13,14}, and a similar 'pitch centre' may exist in non-human primate auditory cortex. In this study, we searched for pitch-selective neurons in the auditory cortex of the common marmoset (*Callithrix jacchus*): a New World primate species sharing a similar hearing range with humans¹⁵. Using single-unit extracellular recordings (see Methods), we found a restricted cortical region near the anterolateral low-

frequency border of AI in the marmoset containing neurons that respond significantly to both pure tones and MFs with similar pitches.

In order for a neuron to be considered pitch-selective, we required that it satisfy two criteria. First, the neuron had to respond significantly to both pure tones and MFs with a similar pitch. Second, all of the harmonics of the MF had to be outside the neuron's excitatory-frequency response area. An example of a neuron's response to acoustic stimuli to test these criteria is shown in Fig. 1 (see also Supplementary Fig. 1). A total of 53 neurons from three marmosets met our criteria for pitch-selectivity. Fifty-one of these neurons were located within a restricted low-frequency region near the anterolateral border of AI and neighboured by the low-frequency regions of R (rostral field) and laterally situated non-primary areas (Fig. 2a, Supplementary Fig. 2a–c). These pitch-selective neurons accounted for 39% (51/131) of the neurons recorded in this region that responded to pure tones. Pitch-selective and non-pitch neurons in this area spanned a similar range of characteristic frequencies (Fig. 2b). Owing to recording time constraints, we initially searched for MF responses using fundamental frequencies near the neuron's characteristic frequency (determined by pure tone). In some pitch-selective neurons, we systematically varied an MF's f_0 in order to determine the neuron's best fundamental frequency. In general, pitch-selective neurons were similarly tuned for their peak responses to pure tones and MFs (Fig. 3b) and always overlapped in their frequency and fundamental frequency tuning for pure tone and MF responses, respectively (Fig. 3a, Supplementary Fig. 3a, b). We did not have any evidence from our experiments to support the existence of neurons with MF and pure tone responses that failed to overlap along the frequency axis. An additional 50 neurons in this region were encountered that did not respond significantly to pure tones, but did respond to narrowband or wideband stimuli such as harmonic complex tones, sinusoidally amplitude- or frequency-modulated tones (sAM, sFM), click trains, or band-pass noise. A subset of these neurons ($n = 10$) only responded to harmonic complex and sAM tones with repetition rates similar in frequency to the characteristic frequencies of neighbouring neurons. These neurons may play a role in processing the pitch of complex sounds; however, they were not included in our analysis of pitch-selective neurons due to an insufficient sample size.

Once we characterized neurons as pitch-selective, we further tested these cells with a variety of complex sounds whose pitch salience were parametrically varied. A click train (see Methods) has a pitch corresponding to its average repetition rate and a pitch salience determined by the regularity of the time intervals between successive clicks. When the timing of individual clicks is 'jittered' to create an irregular click train, the pitch salience decreases with increasing irregularity¹⁶. We tested the effect of a click train's temporal irregularity on neuronal responses in a subset of pitch-selective neurons and found an overall decrease in their discharge rates (Fig. 4a,

¹Laboratory of Auditory Neurophysiology, Department of Biomedical Engineering, Johns Hopkins University School of Medicine, Baltimore, Maryland 21025, USA.

Supplementary Fig. 4a, b). For another subset of pitch-selective neurons, we tested their sensitivity to pitch salience using iterated-ripple-noise (IRN) stimuli, which were constructed by adding broadband noise iteratively to itself with a constant delay¹⁷. Because each iteration of this delay-and-add process increases the temporal regularity of the resulting sound, the pitch strength of the stimulus also increases. Overall, pitch-selective neurons increased their discharge rate as the strength of pitch in the IRN also increased (Fig. 4b, Supplementary Fig. 4c).

Pitch salience is also dependent on the harmonic composition of an MF. Data from studies in humans indicate that the salience of pitch is greater in MFs composed of lower-order harmonics rather than those composed of higher-order harmonics¹. It is the third to fifth harmonics of a harmonic complex tone that contribute the most to its pitch¹. However, it is unknown if this is also the case in marmosets. We observed that pitch-selective neurons usually responded most strongly to harmonic complex sounds containing lower-order harmonics (first to sixth harmonics) (Fig. 4c).

Several important distinctions must be made between this study and previous reports of a neural representation of periodicity in the auditory cortex of the gerbil^{18,19}. In the present study, the pitch-selective neurons had characteristic frequencies that were mostly below 800 Hz (Fig. 2b) and, given the correspondence between characteristic frequency and preferred missing fundamental frequency (Fig. 3b), this closely matches the human perceptual limit of missing fundamental pitch²⁰. Responses at higher best-modulation frequencies (2–3 kHz) were observed in previous studies investigating periodicity-encoding in gerbil auditory cortex^{18,19}. Another difference between these studies was the frequency range of harmonics to which neurons responded. In our study, only MFs

containing harmonics below ~5 kHz evoked significant responses in most pitch-selective neurons (Fig. 4d). This matches the upper frequency limit of an MF's harmonics for its pitch to be perceivable by humans¹. In contrast, the carrier frequencies of sAM tones used in previous studies investigating periodicity responses in gerbil auditory cortex^{18,19} were above 5 kHz. Finally, a crucial distinction between the present study and previous work was the sound level at which MF and sAM acoustic stimuli were delivered, respectively. When the ear is stimulated with two tones (f_1 and f_2), combination tones ($2f_1 - f_2$, $f_2 - f_1$, and so on) are generated by the non-linear mechanics of the cochlea¹. Psychophysical studies show that missing MFs with two components generate combination tones at the f_0 that are 20–25 dB lower than the sound level of individual components²¹. The magnitude of this combination tone increases by 3 dB for every doubling of the number of components. Physiological studies in the inferior colliculus of guinea pigs²² suggest that combination tones at the f_0 can be produced in the range of 17–34 dB below the sound level of the carrier of an amplitude-modulated tone. To avoid the confound of neural responses evoked by combination tones, we strictly limited the sound level of the individual components of MFs used in our experiments to be no more than 10 dB above the neuron's tone response threshold at its characteristic frequency. The outer ear provides an additional amplification to the harmonics of the MF and may affect our estimation of the sound level of combination tones. Although the spectral-specific gain of the outer ear has not been measured in the marmoset, other animal models indicate that the gain increases with frequency (over the frequency range 100–5,000 Hz) with a maximum relative gain between high frequencies and low frequencies of about 10 dB. More than 75% of the pitch-selective neurons (40/53) responded significantly to an MF when the

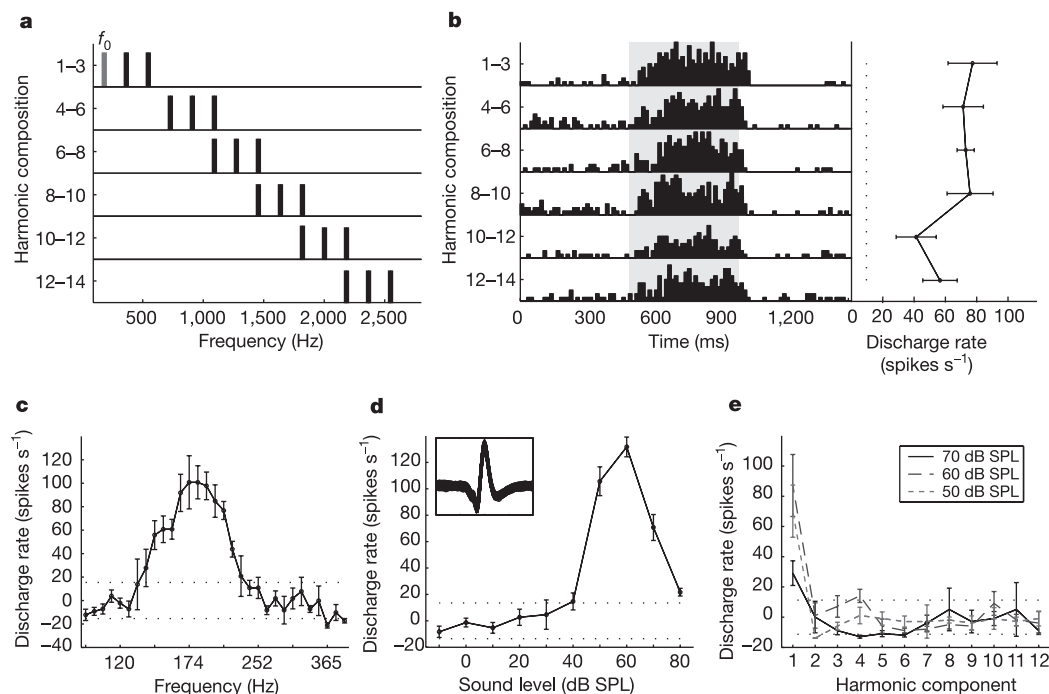


Figure 1 | An example of a pitch-selective neuron (unit M36n-532). Error bars represent standard error of the mean (s.e.m.). The dotted black lines indicate the significance level for discharge rate (± 2 standard deviations away from the spontaneous discharge rate). **a**, Frequency spectra of a series of harmonic complex stimuli. The fundamental frequency component (f_0) and its higher harmonics have equal amplitudes of 50 dB SPL. **b**, Peristimulus time histogram (left) and tuning curve (right) of the neuron's response to the stimuli in **a**. Stimuli were presented from 500 to 1,000 ms (indicated by the shaded region on the left plot). **c**, Frequency

tuning of the neuron derived from pure tones. **d**, Response of the neuron to a pure tone at characteristic frequency (182 Hz) across sound levels (rate-level function). Inset plot shows an overlay of 2,434 digitized action potentials recorded from this neuron (displayed within a 2 ms window). **e**, The neuron's responses to individual harmonics (number 1–12) at three sound levels, respectively. All the harmonics above the f_0 component (first harmonic) were outside the neuron's excitatory frequency response area, and did not elicit significant responses. SPL, sound pressure level.

individual components were set at the neuron's pure tone sound level threshold at its characteristic frequency (Fig. 5a); a situation where combination tones at the neuron's characteristic frequency would be at least 20 dB below its response threshold (or 10 dB assuming the maximum outer-ear differential gain between f_0 and the harmonics of the MF) as estimated by previous studies^{21,22}. As such, the procedures implemented in the present study ensure that the MF responses reported here are not the result of combination tones. Previous studies^{18,19} employed sAM tones delivered at 30 dB or more above a neuron's sound level threshold, making the interpretation of the reported periodicity representation difficult.

Combination tones can be perceptually masked by spectrally overlapping band-pass noise¹. We compared responses to MFs with and without a noise masker for a subset of pitch-selective neurons ($n = 20$). The masker was generated using 1–2 octave band-pass noise centred at the f_0 of the MF and at a sound level -10 to $+10$ dB

relative to the levels of individual components of the MF. None of the pitch-selective neurons studied failed to respond significantly in the presence of the noise masker (Fig. 5b). The approximate 50:50 ratio of neurons whose discharge rates increased or decreased in the presence of the noise masker may be due to the proximity of this cortical pitch area to both the core and belt regions of auditory cortex that show preferences for tonal or noisy sounds, respectively²³. Less than half of the neurons from Fig. 5b that were tested responded significantly to the noise masker when it was played alone (Supplementary Fig. 5a, b).

Magnetoencephalography studies in humans suggest both a parallel²⁴ and orthogonal²⁵ topographical organization of pitch relative to the cochleotopic map in AI. In addition, a recent optical imaging study in gerbils¹⁹ has suggested a horseshoe-shaped topographical map for periodicity that is superimposed on a linear cochleotopic map. Due to the small size of the cortical area containing pitch-selective neurons ($\sim 1 \text{ mm}^2$) (Fig. 2a, Supplementary Fig. 2a, b), we could not determine any topographical arrangement of best pitch encoded by these neurons. Pitch-selective and non-pitch neurons within this region had characteristic frequencies spanning the same frequency range (Fig. 2b). However, given that non-pitch neurons encoding low frequencies are present in the same region of auditory cortex, these data support a parallel topographical representation of pitch and frequency. The two characteristic frequency distributions were significantly different ($P = 0.0251$, Wilcoxon rank-sum test) with pitch-selective neurons biased towards lower-frequency characteristic frequencies; however, bandwidth and peak latency were not significantly different between these two groups of neurons. While the range of characteristic frequencies encountered from pitch-selective neurons fell below the f_0 of most marmoset vocalizations

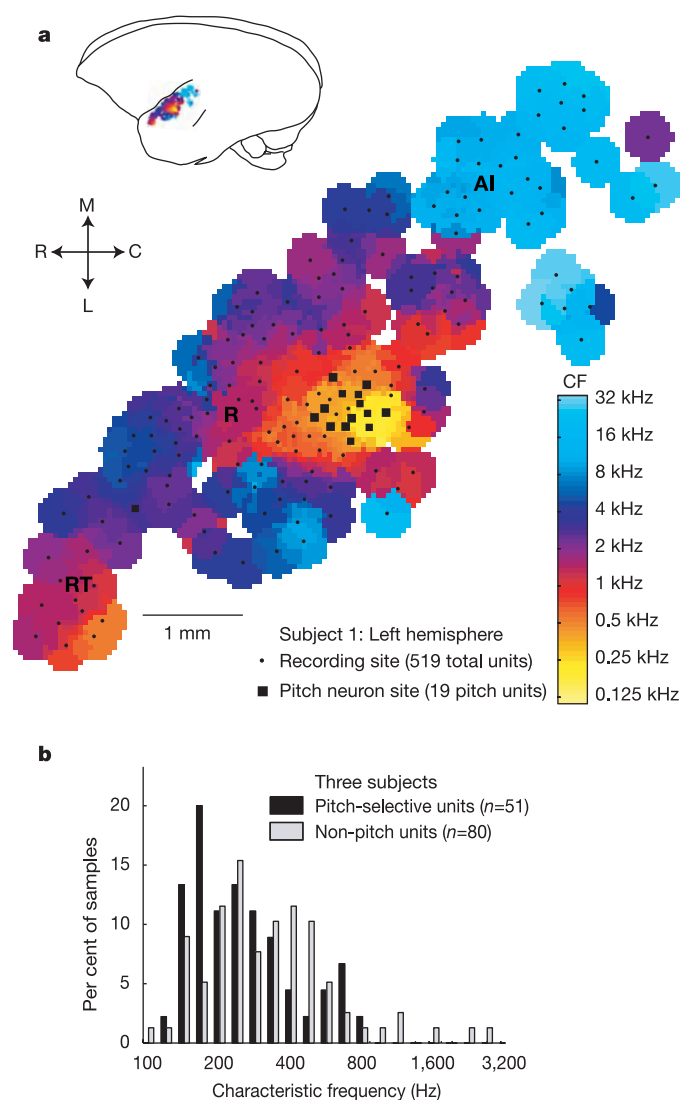


Figure 2 | Location and characteristic frequency distribution of the pitch area in marmoset auditory cortex. a, Characteristic frequency topographical map from the left hemisphere of one marmoset. Pitch-selective neurons (black squares) were found clustered near the anterolateral border of AI. Frequency reversals indicate the borders between AI/R and R/RT (rostral temporal field). **b**, The characteristic frequency distribution from pitch-selective and non-pitch neurons within the pitch area of three marmosets. M, medial; C, caudal; L, lateral; R, rostral; CF, characteristic frequency.

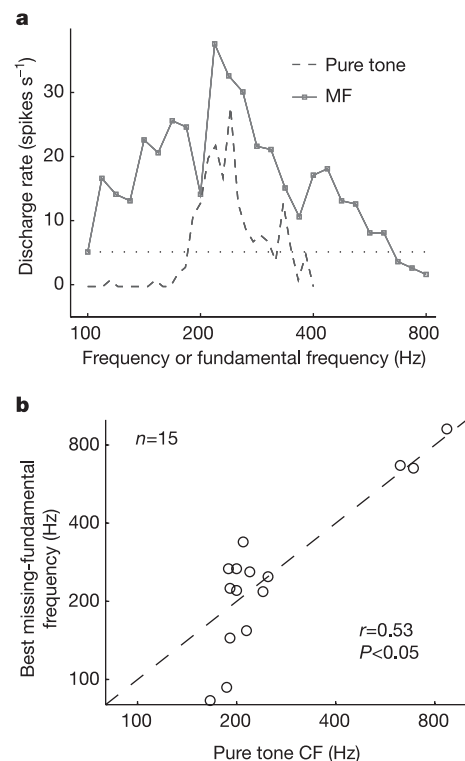


Figure 3 | Pitch-selective neurons share a similar tuning for pure tones and MFs. a, An example of an individual pitch-selective neuron's tuning to pure tone frequency and the fundamental frequency of MFs respectively. (unit M2p-201) **b**, A comparison of the characteristic frequency and the best missing-fundamental frequency responses from 15 pitch-selective neurons. The Spearman correlation coefficient (r) is displayed on the plot and is statistically significant ($P < 0.05$).

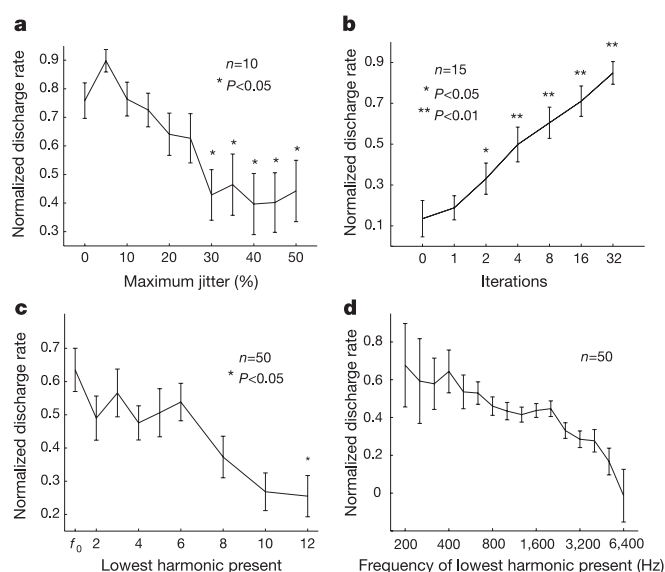


Figure 4 | Pitch-selective neurons are sensitive to pitch salience. Error bars represent s.e.m. Statistical significance was determined using Wilcoxon rank-sum test. Responses were normalized by the maximum response elicited within the stimulus set. **a**, Averaged population response of pitch-selective neurons to irregular click trains as a function of maximum jitter. The response to a regular click train was used as a reference for statistical comparison at other jitter values. **b**, Averaged population response as a function of the iterations of IRN stimuli. The response to IRN stimuli with 0 iterations was used as a reference for statistical comparison at other iterations. **c**, Averaged population response as a function of the lowest harmonic presented in the MF stimuli. The reference for statistical comparison was harmonic complex sounds with their fundamental frequency present. **d**, Averaged population response as a function of the frequency of the lowest harmonic presented in the MF stimuli.

(4–8 kHz), marmosets produce several call types (for example, ‘egg’ call, $f_0 \approx 800$ Hz) that have fundamental frequencies near the upper range of the characteristic frequencies of pitch-selective neurons²⁶. It is important to note that marmosets hear sounds containing harmonic structure from other animals and the environment in their natural habitat.

The cortical region containing pitch-selective neurons appears to be on the border of core areas AI and R, and lateral belt areas AL (anterolateral) and ML (middle lateral), without spanning the entire tonotopic representation of any one of these four areas. This may be a frequency-specific and functionally specialized area of auditory cortex in primates, analogous to areas of auditory cortex of the mustached bat (*Pteronotus parnellii*) that contain combination-sensitive neurons²⁷.

Lower-order harmonics of a complex tone are resolved by the auditory system, and the estimates of the frequencies of these components can be used to determine the pitch²⁸. However, when the harmonics of a complex tone are not resolved by the auditory system, only the temporal information (repetition rate) of the acoustic waveform can be used to determine the pitch²⁹. How marmosets perceive these MFs and, more specifically, to what extent they use spectral and temporal pitch mechanisms remains to be studied in future behavioural and physiological experiments. Given that the size of the cochlea is smaller in marmosets than in humans, it is probable that some of the lower-order harmonics resolved in the human are unresolved in the marmoset. As such, the MF responses that we observed were most probably evoked by both resolved and unresolved harmonics. Spectral and temporal processing strategies may ultimately be unified in auditory cortex, providing a single central neural correlate for the perception of pitch.

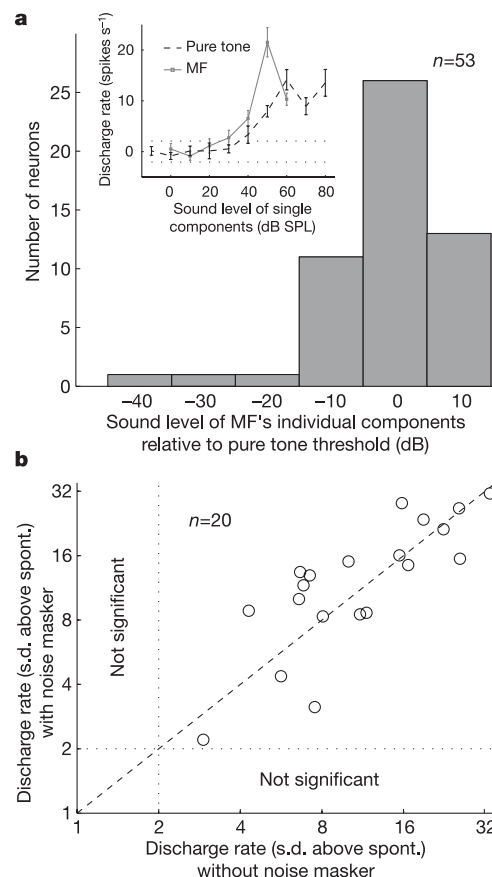


Figure 5 | MF responses are not caused by combination tones.

a, Distribution of sound level threshold for individual components of the MF response relative to the sound level threshold for a pure tone response at the neuron's characteristic frequency. Inset plot shows rate-level functions from a pitch-selective neuron (unit M41o-294) for pure tones and MFs. The two dotted lines indicate two standard deviations from the spontaneous discharge rate. Error bars represent s.e.m. **b**, Scatter plot comparing responses to MFs with and without the presence of a noise masker for 20 pitch-selective neurons. All the neurons tested had significant discharge rates for both conditions. The two dotted lines parallel to the axes indicate two standard deviations (s.d.) from the spontaneous discharge rate. The diagonal line has a slope of 1.

METHODS

Animal preparation and recording. Details of experimental procedures can be found in recent publications from our laboratory³⁰. Single-unit recordings were conducted in awake marmosets (subjects 1–3: M2p (left hemisphere), M36n (right hemisphere), M41o (left hemisphere)) sitting quietly in a semi-restraint device with their head immobilized, within a double-walled soundproof chamber (Industrial Acoustics) whose interior is covered by 3-inch acoustic absorption foam (Sonex). Because the auditory cortex of the marmoset lies largely on the lateral surface of the temporal lobe, high-impedance tungsten microelectrodes (3–5 M Ω) could be inserted perpendicular to the cortical surface. Electrodes were mounted on a micromanipulator (Narishige) and advanced by a manual hydraulic microdrive (Trent Wells). Action potentials were detected on-line using a template-based spike sorter (Multi-Spike Detector; Alpha Omega Engineering) and continuously monitored by the experimenter while data recording progressed. Typically 5–15 electrode penetrations were made within a miniature recording hole (diameter ~ 1 mm), after which the hole was sealed with dental cement and another hole opened for new electrode penetrations. Neurons were recorded from all cortical layers, but most commonly from supragranular layers.

Generation of acoustic stimuli. Acoustic stimuli were generated digitally and delivered by a free-field loudspeaker located one metre directly in front of the animal. All sound stimuli were generated at a 100 kHz sampling rate and

low-pass filtered at 50 kHz. Harmonic artifacts were at least 43 dB lower than the fundamental at 80 dB SPL (sound pressure level). The difference grew as the sound level of the fundamental decreased. The sound level of individual frequency components used in this study was no higher than 80 dB SPL.

Frequency tuning curves and rate-level functions were generated using pure-tone stimuli of 200 ms in duration with interstimulus intervals of >500 ms, and had a minimum of 5 repetitions. MF, IRN, and click-train stimuli were 500 ms in duration with intertrial intervals of at least 1 s, and had a minimum of 10 repetitions. All stimuli were presented in a randomly shuffled order. Pure-tone stimuli intensity levels were generally 10–20 dB above threshold for neurons with monotonic rate-level functions, or at preferred levels for non-monotonic neurons. Harmonic complex tones were composed of 3 or 9 components in either cosine or Schroeder negative phase. The individual components of all harmonic complex tone stimuli were presented at no more than 10 dB above the neuron's sound level threshold at its characteristic frequency. Components of the MF were considered outside the neuron's excitatory frequency response area if each component, when played individually at 0, +10 and +20 dB relative to its sound level within the harmonic complex, did not evoke a significant response. Sound levels were varied in 10 dB steps.

Noise maskers were typically 1–2 octave band-pass noise centred at the missing fundamental frequency (near the unit's characteristic frequency). The sound level of the noise masker ranged from +10 to –10 dB relative to the individual harmonics. Noise maskers were played simultaneously with MFs.

Regular click trains had inter-click intervals equal to $1/f_0$ where f_0 was the preferred fundamental frequency of the neuron. Rectangular clicks (broadband) or narrowband clicks made of brief pulses of white noise or a tone (at an integer multiple of the f_0) were used to generate click trains. Rectangular click trains had a width of 0.1 ms while narrowband clicks³⁰ had each pulse windowed by a gaussian envelope with a sigma of 0.1–0.4. An irregular click train was constructed by shifting each click of a regular click train relative to a previous click by an amount of time proportional to the ISI and randomly selected from a uniform distribution $S_x = [-J, J]$, where J equals the maximum possible jitter. The maximum jitter in the irregular click train stimulus set was varied between 5 to 50%.

Generation of cortical characteristic frequency maps. Single units with significant neuronal discharges to tones, band-pass noise, or other narrowband stimuli (for example, sAM, sFM) were used to generate cortical characteristic frequency maps. The characteristic frequency of each location on the map is determined by the median characteristic frequency of all electrode tracks within 0.25 mm. Electrode track characteristic frequencies were calculated by computing the median characteristic frequency of units within the track.

Data analysis. The mean spontaneous discharge rate was subtracted during the calculation of a neuron's mean driven discharge rate over the entire duration of the stimulus. Mean driven discharge rates greater than 2 standard deviations above the spontaneous discharge rate were considered significant. The peak MF response from every pitch-selective neuron was also determined to be significant ($P < 0.05$) using a Wilcoxon rank-sum test.

Received 12 January; accepted 26 May 2005.

- Moore, B. C. J. *An Introduction to the Psychology of Hearing* (Academic, London, 2003).
- Tomlinson, R. W. & Schwarz, D. W. Perception of the missing fundamental in nonhuman primates. *J. Acoust. Soc. Am.* **84**, 560–565 (1988).
- Heffner, H. & Whitfield, I. C. Perception of the missing fundamental by cats. *J. Acoust. Soc. Am.* **59**, 915–919 (1976).
- Patterson, R. D., Uppenkamp, S., Johnsrude, I. S. & Griffiths, T. D. The processing of temporal pitch and melody information in auditory cortex. *Neuron* **36**, 767–776 (2002).
- Penagos, H., Melcher, J. R. & Oxenham, A. J. A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. *J. Neurosci.* **24**, 6810–6815 (2004).
- Cariani, P. A. & Delgutte, B. Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *J. Neurophysiol.* **76**, 1698–1716 (1996).
- Cedolin, L. & Delgutte, B. Pitch of complex tones: rate-place and interspike interval representation in the auditory nerve. *J. Neurophysiol.* **94**, 347–362 (2005).
- Whitfield, I. C. Auditory cortex and the pitch of complex tones. *J. Acoust. Soc. Am.* **67**, 644–647 (1980).
- Zatorre, R. J. Pitch perception of complex tones and human temporal-lobe function. *J. Acoust. Soc. Am.* **84**, 566–572 (1988).
- Warrier, C. M. & Zatorre, R. J. Right temporal cortex is critical for utilization of melodic contextual cues in a pitch constancy task. *Brain* **127**, 1616–1625 (2004).
- Schwarz, D. W. & Tomlinson, R. W. Spectral response patterns of auditory cortex neurons to harmonic complex tones in alert monkey (*Macaca mulatta*). *J. Neurophysiol.* **64**, 282–298 (1990).
- Fishman, Y. I., Reser, D. H., Arezzo, J. C. & Steinschneider, M. Pitch vs. spectral encoding of harmonic complex tones in primary auditory cortex of the awake monkey. *Brain Res.* **786**, 18–30 (1998).
- Morel, A. & Kaas, J. H. Subdivisions and connections of auditory cortex in owl monkeys. *J. Comp. Neurol.* **318**, 27–63 (1992).
- Morel, A., Garraghty, P. E. & Kaas, J. H. Tonotopic organization, architectonic fields, and connections of auditory cortex in macaque monkeys. *J. Comp. Neurol.* **335**, 437–459 (1993).
- Fay, R. R. *Hearing in Vertebrates: A Psychophysics Databook* (Hill-Fay, Winnetka, 1988).
- Pollack, I. Detection and relative discrimination of auditory “jitter”. *J. Acoust. Soc. Am.* **43**, 308–315 (1968).
- Yost, W. A., Patterson, R. & Sheft, S. The role of the envelope in processing iterated rippled noise. *J. Acoust. Soc. Am.* **104**, 2349–2361 (1998).
- Schulze, H. & Langner, G. Periodicity coding in the primary auditory cortex of the Mongolian gerbil (*Meriones unguiculatus*): two different coding strategies for pitch and rhythm? *J. Comp. Physiol. A* **181**, 651–663 (1997).
- Schulze, H., Hess, A., Ohl, F. W. & Scheich, H. Superposition of horseshoe-like periodicity and linear tonotopic maps in auditory cortex of the Mongolian gerbil. *Eur. J. Neurosci.* **15**, 1077–1084 (2002).
- Ritsma, R. J. Existence region of the tonal residue. I. *J. Acoust. Soc. Am.* **34**, 1224–1229 (1962).
- Pressnitzer, D. & Patterson, R. D. Distortion products and the perceived pitch of harmonic complex tones. in *Physiological and Psychophysical Bases of Auditory Function* (eds Breebart, D. J., Houtsma, A. J. M., Kohlrausch, A., Pries, V. F. & Schoonen, R.) 97–104 (Shaker, Maastricht, 2001).
- McAlpine, D. Neural sensitivity to periodicity in the inferior colliculus: evidence for the role of cochlear distortions. *J. Neurophysiol.* **92**, 1295–1311 (2004).
- Rauschecker, J. P. & Tian, B. Processing of band-passed noise in the lateral auditory belt cortex of the rhesus monkey. *J. Neurophysiol.* **91**, 2578–2589 (2004).
- Pantev, C., Hoke, M., Lutkenhoner, B. & Lehnertz, K. Tonotopic organization of the auditory cortex: pitch versus frequency representation. *Science* **246**, 486–488 (1989).
- Langner, G., Sams, M., Heil, P. & Schulze, H. Frequency and periodicity are represented in orthogonal maps in the human auditory cortex: evidence from magnetoencephalography. *J. Comp. Physiol. A* **181**, 665–676 (1997).
- Epple, G. Comparative studies on vocalization in marmoset monkeys (*Haplorhina*). *Folia Primatol. (Basel)* **8**, 40 (1968).
- Suga, N. Processing of auditory information carried by species-specific complex sounds. in *The Cognitive Neurosciences* (ed. Gazzanica, M. S.) 295–313 (MIT Press, Cambridge, Massachusetts, 1994).
- Goldstein, J. L. An optimum processor theory for the central formation of the pitch of complex tones. *J. Acoust. Soc. Am.* **54**, 1496–1516 (1973).
- Schouten, J. F. The residue and the mechanism of hearing. *Proc. K. Ned. Akad. Wet.* **43**, 991–999 (1940).
- Lu, T., Liang, L. & Wang, X. Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. *Nature Neurosci.* **4**, 1131–1138 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by NIH grants to X.W. and D.B. We thank B. Delgutte, D. McAlpine, E. Young, B. Moore and members of the Laboratory of Auditory Neurophysiology for their comments and suggestions related to this manuscript, and A. Pistorio, E. Bartlett and E. Issa for assistance with animal care. E. Issa contributed data to the characteristic frequency maps.

Author Contributions D.B. and X.W. designed the experiment and co-wrote the paper. D.B. carried out the electrophysiological recordings and data analysis.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.B. (dbendor@bme.jhu.edu) or X.W. (xwang@bme.jhu.edu).

LETTERS

Translational control of hippocampal synaptic plasticity and memory by the eIF2 α kinase GCN2

Mauro Costa-Mattioli¹, Delphine Gobert⁵, Heather Harding⁶, Barbara Herdy¹, Mounia Azzi⁵, Martin Bruno², Michael Bidinosti¹, Cyrinne Ben Mamou³, Edwige Marcinkiewicz⁷, Madoka Yoshida¹, Hiroaki Imataka⁸, A. Claudio Cuello², Nabil Seidah⁷, Wayne Sossin⁴, Jean-Claude Lacaille⁵, David Ron⁶, Karim Nader³ & Nahum Sonenberg¹

Studies on various forms of synaptic plasticity have shown a link between messenger RNA translation, learning and memory. Like memory, synaptic plasticity includes an early phase that depends on modification of pre-existing proteins, and a late phase that requires transcription and synthesis of new proteins^{1,2}. Activation of postsynaptic targets seems to trigger the transcription of plasticity-related genes. The new mRNAs are either translated in the soma or transported to synapses before translation. GCN2, a key protein kinase, regulates the initiation of translation. Here we report a unique feature of hippocampal slices from *GCN2*^{-/-} mice: in CA1, a single 100-Hz train induces a strong and sustained long-term potentiation (late LTP or L-LTP), which is dependent on transcription and translation. In contrast, stimulation that elicits L-LTP in wild-type slices, such as four 100-Hz trains or forskolin, fails to evoke L-LTP in *GCN2*^{-/-} slices. This aberrant synaptic plasticity is mirrored in the behaviour of *GCN2*^{-/-} mice in the Morris water maze: after weak training, their spatial memory is enhanced, but it is impaired after more intense training. Activated GCN2 stimulates mRNA translation of ATF4, an antagonist of cyclic-AMP-response-element-binding protein (CREB). Thus, in the hippocampus of *GCN2*^{-/-} mice, the expression of ATF4 is reduced and CREB activity is increased. Our study provides genetic, physiological, behavioural and molecular evidence that GCN2 regulates synaptic plasticity, as well as learning and memory, through modulation of the ATF4/CREB pathway.

Translation of eukaryotic mRNAs is regulated primarily at the level of initiation³. Binding of the initiator tRNA, Met-tRNA^{Met}, to the 40S subunit is facilitated by the initiation factor 2 (eIF2) which forms a ternary complex with GTP and Met-tRNA^{Met}. Although phosphorylation of the α subunit of eIF2 can inhibit general translation^{4,5}, it stimulates the mRNA translation of the transcriptional modulator ATF4 (ref. 6), which inhibits synaptic plasticity and behavioural learning in various phyla^{7–10}. In view of the need for translation for the modulation of synaptic activity and strong evidence that phosphorylation of eIF2 α controls translation of ATF4 mRNA^{6,11,12}, eIF2 α kinase(s) may regulate synaptic plasticity. Because GCN2 is the most evolutionarily conserved eIF2 α kinase and GCN2 mRNA is enriched in the brain of flies¹³ and mammals (as well as in liver)^{14,15} (see Supplementary Fig. 1), we explored the role of GCN2 in synaptic plasticity and behavioural learning.

The GCN2 gene was inactivated by homologous recombination in embryonic stem cells (Supplementary Fig. 1A and Supplementary

Information). Hippocampal immunohistochemistry and *in situ* hybridization show that GCN2, normally expressed mainly in CA1 and CA3 and also in dentate gyrus, is undetectable in brain slices from *GCN2*^{-/-} mice (Supplementary Figs 1 and 2).

There were no gross morphological changes in the hippocampus or other regions of the brain of *GCN2*^{-/-} mice (Supplementary Fig. 3), and basal synaptic transmission in CA1 was unaltered as indicated by the following: first, the relation of fEPSPs to stimulus intensity; second, the size of the fibre volley; third, paired-pulse facilitation (PPF); and fourth, peak response to tetanic stimulation (Supplementary Fig. 4 and Supplementary Information). Normally, a single high-frequency tetanus (100 Hz for 1 s) elicits in the Schaffer collateral/commissural pathway a transient form of long-term potentiation known as early LTP (E-LTP), which decays in 2–3 h and does not require RNA or protein synthesis¹. In slices from *GCN2*^{-/-} mice, a single tetanus induces a robust and sustained L-LTP (Fig. 1a; at 180 min, $P < 0.001$) and the initial potentiation is greater than in slices from wild-type (WT) mice (Fig. 1a; at 15 min, $P < 0.05$). This increase was synapse-specific because a control input that received only test stimulation remained stable for the entire experimental session (Supplementary Fig. 5A). GCN2 therefore affects the duration of LTP and its initial amplitude.

Like the L-LTP normally elicited by four tetanic trains, the L-LTP induced by a single tetanus in slices from *GCN2*^{-/-} mice depends on cAMP-dependent protein kinase (PKA) (Fig. 1b; at 180 min, $P < 0.01$), new mRNA (Fig. 1b; at 180 min, $P < 0.01$) and protein synthesis (Fig. 1b; at 180 min, $P < 0.001$) and is resistant to depotentiation (Supplementary Fig. 5C). As expected, E-LTP elicited in slices from WT mice by a single tetanus was not affected by inhibiting these pathways and could be depotentiated (Supplementary Fig. 5B, C). Anisomycin (a translation inhibitor) and actinomycin D (a transcription inhibitor) not only prevented the persistence of LTP in slices from *GCN2*^{-/-} mice, but also caused an immediate decrease in the early phase of LTP (Fig. 1b). Similarly to our results, the effects of anisomycin on L-LTP in the Schaffer collateral pathway often show an immediate decrement in the magnitude of potentiation, indicating that protein-synthesis-dependent processes are required early after L-LTP induction^{16,17}. The early effect of actinomycin D indicates that the increased amplitude of initial potentiation might be due to the translation of immediate-early genes whose mRNAs are quickly turned over. Indeed, when actinomycin D is applied 15 min (instead of 30 min)

¹Department of Biochemistry and McGill Cancer Center, ²Department of Pharmacology and Therapeutics, ³Department of Psychology, ⁴Department of Neurology and Neurosurgery, McGill University, Montreal, Quebec H3G 1Y6, Canada. ⁵Département de physiologie, Centre de Recherche en Sciences Neurologiques, Université de Montréal, C.P. 6128, Succ. Centre-Ville, Montréal, Québec H3C 3J7, Canada. ⁶Skirball Institute, Departments of Medicine, Cell Biology and Pharmacology, NYU School of Medicine, New York, New York 10016, USA. ⁷Laboratory of Biochemical Neuroendocrinology, Clinical Research Institute of Montreal, 110 Pine Avenue, West Montreal, Quebec H2W 1R7, Canada. ⁸Genomic Sciences Center, RIKEN, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan.

before the onset of tetanization to minimize the effects of steady-state levels of rapidly turning-over mRNAs, the drug did not have an immediate effect (Fig. 1c; at 60 min, $P > 0.05$). Instead there was a delayed decrease in LTP, which was consistent with the lack of induction of new mRNAs necessary for the maintenance of LTP (Fig. 1c; at 180 min, $P < 0.01$).

According to these observations, deletion of GCN2 leads to an enhanced response to a single tetanus, resulting in L-LTP instead of E-LTP. Does GCN2 deletion also affect the L-LTP normally induced by repeated tetani? To address this question we examined L-LTP induced in CA1 by two different protocols: tetanic stimulation with four trains at 100 Hz, and forskolin, an activator of PKA¹⁸. As expected, in slices from WT mice, four trains elicited L-LTP that persisted for at least 4 h. By contrast, in slices from *GCN2*^{-/-} mice, the LTP decayed to baseline within 3 h (Fig. 1d; at 240 min, $P < 0.01$). In slices from WT mice, forskolin elicited the usual L-LTP whereas in *GCN2*^{-/-} slices the L-LTP was not sustained (Supplementary Fig. 5D). The GCN2 deletion specifically affected LTP because long-term depression (LTD), which is induced by low-frequency stimulation or by incubation with an agonist of group I mGluRs, 3,5-dihydroxyphenylglycine (DHPG)¹⁹, was unaltered in *GCN2*^{-/-} slices (Supplementary Fig. 6 and Supplementary Information).

Activation of GCN2 can inhibit the initiation of translation by eIF2 α phosphorylation but, paradoxically, it stimulates the translation of ATF4 mRNA⁶. We therefore measured eIF2 α phosphorylation in hippocampal extracts from WT and *GCN2*^{-/-} mice and found that it was lower ($50 \pm 19\%$) in *GCN2*^{-/-} mice (Fig. 2a). Consistent with this finding was the observation that ATF4 mRNA was shifted to the lighter polysome fractions of hippocampal extracts from *GCN2*^{-/-} mice (Fig. 2b, c). In agreement with a weak basal translation of ATF4 mRNA, ATF4 protein was correspondingly lower ($49 \pm 11\%$; Fig. 2e). By contrast, β -actin mRNA sedimented predominantly in the heavy polysome fractions, as would be expected for an efficiently translated mRNA (Fig. 2b, d). Thus, GCN2 deletion leads to a decrease in translation of ATF4 mRNA in the hippocampus. In accordance with the inhibition of CREB by ATF4, decreased translation of ATF4 mRNA in *GCN2*^{-/-} mice was associated with enhanced CREB function: expression of immediate-early genes regulated by CREB (*BDNF*, *c-fos*, *Egr-1*) was 25–35% greater in *GCN2*^{-/-} hippocampal extracts (Fig. 2f).

To further investigate how synaptic plasticity affects GCN2, we examined the effects of forskolin or four trains at 100 Hz (both induce L-LTP and stimulate CRE-mediated gene expression)²⁰ on GCN2 and eIF2 α phosphorylation. Both protocols decreased GCN2 and eIF2 α phosphorylation in WT but not in *GCN2*^{-/-} slices (Fig. 2g

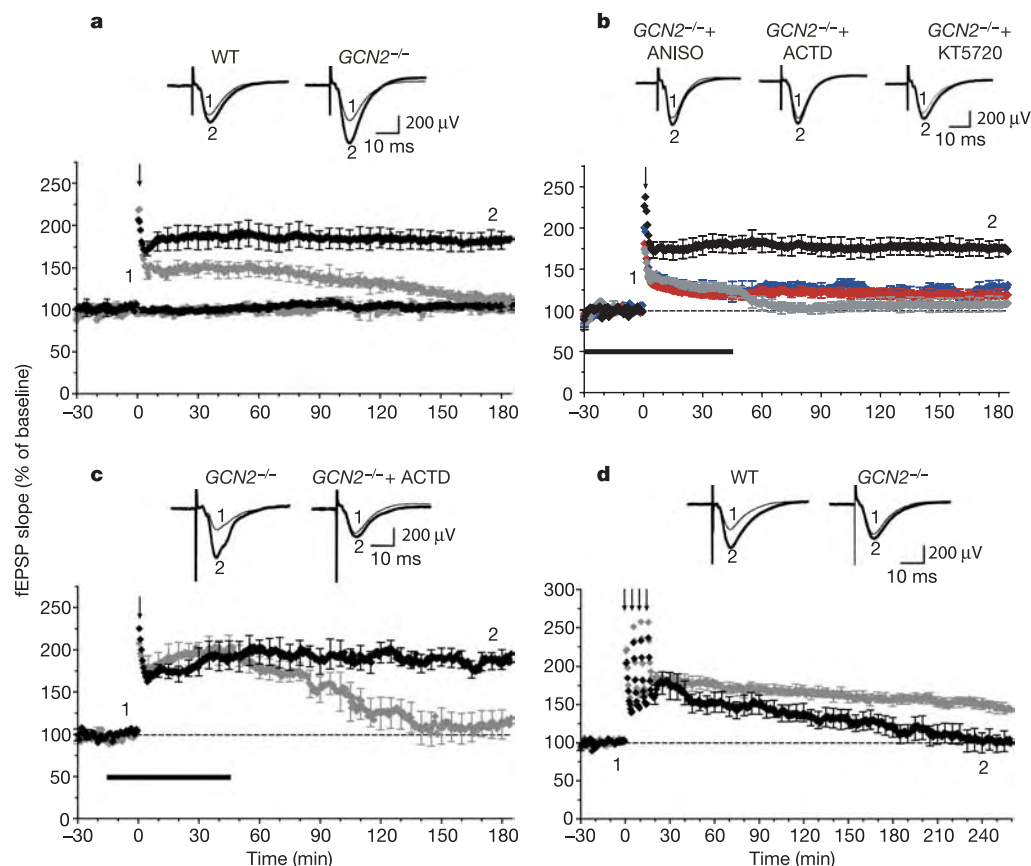


Figure 1 | Unusual properties of LTP induced in slices from *GCN2*^{-/-} mice. **a**, One train (100 Hz for 1 s; vertical arrow) of high-frequency stimulation (HFS) elicited E-LTP in WT slices but produced a robust L-LTP in *GCN2*^{-/-} slices. Black diamonds, *GCN2*^{-/-} ($n = 7$; six mice); grey diamonds, WT ($n = 5$; four mice); black circles, *GCN2*^{-/-} without tetanic stimulation ($n = 6$; four mice). **b**, Sustained LTP in *GCN2*^{-/-} slices is decreased by anisomycin (ANISO, 40 μ M), actinomycin D (ACTD, 40 μ M) or the PKA inhibitor KT5720 (1 μ M), added for the duration of the horizontal bar. Black diamonds, *GCN2*^{-/-} plus vehicle ($n = 9$; eight mice); red, *GCN2*^{-/-} plus

ANISO ($n = 8$; six mice); grey, *GCN2*^{-/-} plus ACTD ($n = 10$; six mice); blue, *GCN2*^{-/-} plus KT5720 ($n = 7$; five mice). **c**, The enhanced LTP in *GCN2*^{-/-} slices is reduced at later time points (>90 min) by 40 μ M ACTD (horizontal bar), when applied 15 min before and 45 min after tetanus. Black diamonds, *GCN2*^{-/-} ($n = 6$; four mice); grey diamonds, *GCN2*^{-/-} plus ACTD ($n = 7$; five mice). **d**, L-LTP induced by four 100 Hz trains at 5 min intervals (vertical arrows) is stable in WT slices (grey diamonds; $n = 10$; six mice) but not in *GCN2*^{-/-} slices (black diamonds; $n = 9$; seven mice). All results show mean $\pm$ s.e.m.

and Supplementary Fig. 7A). However, E-LTP elicited by a single train was not associated with a decrease in GCN2 and eIF2 α phosphorylation (Supplementary Fig. 7B). GCN2 activity is therefore regulated by two forms of strong stimulation that elicit L-LTP, but not by a weaker stimulation that induces only E-LTP.

The effects of GCN2 deletion on long-term learning and memory were first studied in a fear conditioning paradigm. Fear conditioning by two tone-shock pairings has two components. One is contextual fear conditioning, which associates the training context and the footshock and requires both the hippocampus and the amygdala. The second, which associates the tone and the footshock, requires the amygdala but not the hippocampus²¹. When tested 1 and 10 days after training, GCN2^{-/-} mice showed a deficit in contextual fear conditioning (Fig. 3a, $P < 0.05$, and Supplementary Information). By contrast, auditory fear conditioning (tested in a different chamber) was intact (Fig. 3b, $P > 0.05$, and Supplementary Information).

Next, hippocampus-dependent spatial memory was tested in the Morris water maze²². In the course of training (three times a day, at 30-min intervals) the performance of both groups improved (Fig. 4a, $P < 0.001$), but WT mice learned faster than GCN2^{-/-} mice (Fig. 4a; at 5 days, $P < 0.01$). In probe tests performed after the end of training, the platform was removed and the mice were allowed to search for 60 s (Fig. 4b). Unlike WT mice (Fig. 4b, $P < 0.001$), GCN2^{-/-} mice showed no preference for the training quadrant (Fig. 4b, $P > 0.05$) and fewer platform crossings (Fig. 4c, $P < 0.001$). Vision and locomotor functions were equally efficient in WT and GCN2^{-/-} mice, as judged by swimming speed ($P > 0.05$) and

latency of escape to a visible platform ($P > 0.05$). Thus, GCN2 deletion is associated with a specific impairment of hippocampus-dependent learning and memory.

Because a single tetanus elicits L-LTP in slices from GCN2^{-/-} mice (Fig. 1a), we reasoned that mnemonic processes might be enhanced during weaker conditioning. Indeed, when mice were trained only once (compared with three times) a day, Tukey's test showed that escape latencies on day 5 were shorter for GCN2^{-/-} than WT mice (Fig. 4d; at 5 days, $P < 0.02$). Enhanced spatial learning by GCN2^{-/-} mice was also evident in the probe tests that were conducted 3 days after the end of training (Fig. 4e). According to repeated-measures analysis of variance (ANOVA), the GCN2^{-/-} mice spent significantly more time in the target quadrant ('trained' in Fig. 4e) than WT mice did ($P < 0.001$). Thus, in agreement with the findings on LTP, memory is enhanced after weak training.

The major finding of this study is that a decrease in threshold for L-LTP in CA1 (in slices) is associated with an improved spatial memory of weak conditioning in GCN2^{-/-} mice. A switch from short-term to long-term plasticity^{8,10,23,24} is generally associated with enhanced gene expression. Indeed, CREB-dependent gene expression is increased in GCN2^{-/-} mice. Thus, GCN2 could effect long-lasting changes in plasticity by modulating CREB activity. The dependence of the early phase of LTP in GCN2^{-/-} mice on transcription and translation may be due to translation of mRNAs coding for CRE-dependent immediate-early genes (which are upregulated at the basal state). Because they turn over rapidly^{25,26}, these mRNAs are likely to be downregulated during the 30 min of pre-incubation with actinomycin D, whereas in WT slices they are scarce

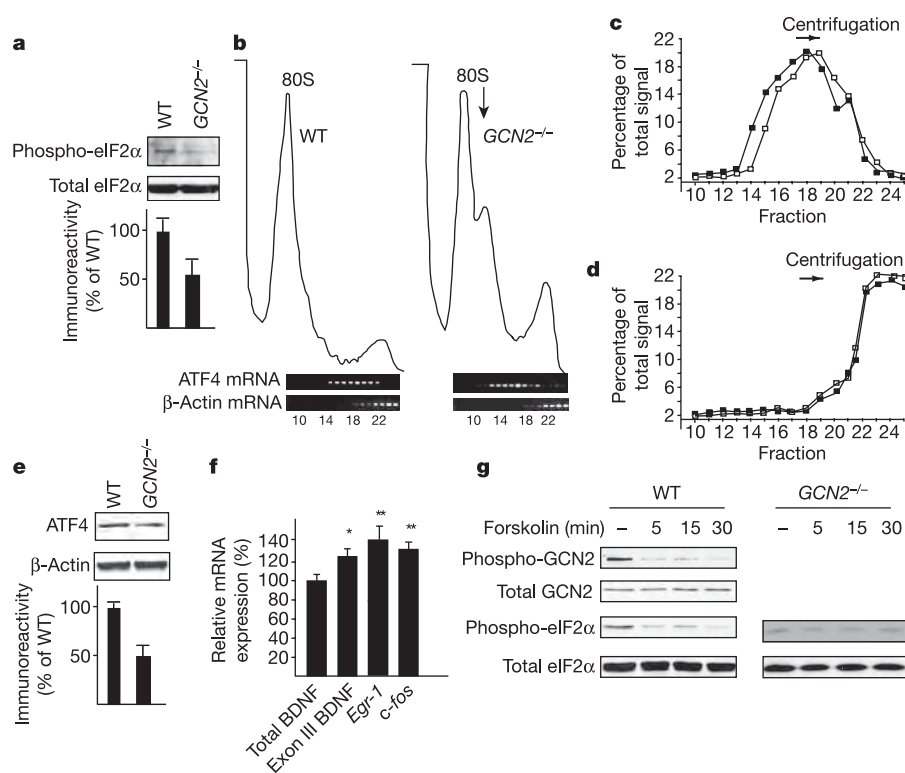


Figure 2 | ATF4 mRNA translation is downregulated in GCN2^{-/-} mice.

a, Western blots performed on hippocampal extracts show that eIF2 α phosphorylation is decreased in GCN2^{-/-} mice ($n = 3$) compared with WT mice ($n = 3$). **b**, In polysome profiles from hippocampal extracts, ATF4 mRNA is in lighter fractions in GCN2^{-/-} (right) than in WT controls (left), as determined by RT-PCR analysis. **c**, Quantification of the band intensities in each fraction from ATF4 mRNA in **b**. Open squares, WT; filled squares, GCN2^{-/-}. **d**, For data in **b**, band intensities are quantified for each fraction of β -actin mRNA. Open squares, WT; filled squares, GCN2^{-/-}. **e**, In pooled

hippocampal extracts, expression of ATF4 is decreased in GCN2^{-/-} mice. **f**, Real-time RT-PCR analysis reveals increased expression of CREB-dependent genes in hippocampal extracts from GCN2^{-/-} compared with that in WT mice (for both, $n = 5$); mRNA expression is given as percentage of controls. Asterisk, $P < 0.05$; two asterisks, $P < 0.01$. Error bars are s.e.m. **g**, Forskolin decreases GCN2 and eIF2 α phosphorylation. In immunoblots of homogenates of CA1 region (from slices frozen immediately after stimulation), phosphorylated GCN2 and eIF2 α are decreased 5 min after application of forskolin.

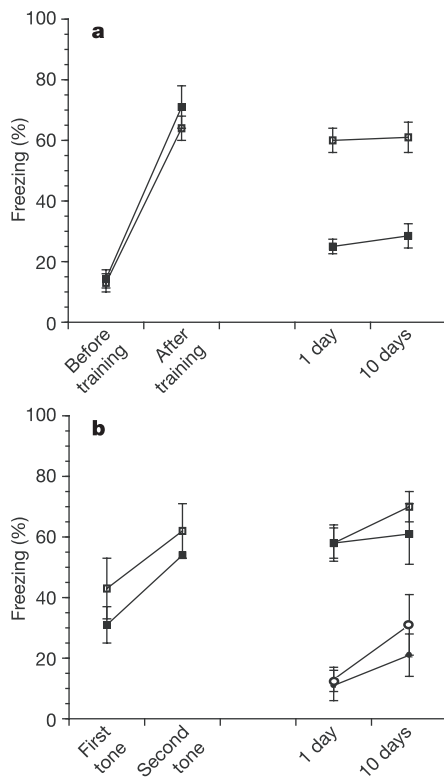


Figure 3 | $GCN2^{-/-}$ mice are impaired in contextual but not auditory fear conditioning. **a**, Acquisition of contextual freezing that compares the 2-min period before the first shock (before training) and 1-min period after the last shock (after training) is similar in $GCN2^{-/-}$ (filled squares, $n = 10$) and WT (open squares, $n = 12$) mice. However, $GCN2^{-/-}$ mice are impaired 1 and 10 days after acquisition. **b**, $GCN2^{-/-}$ mice (filled symbols) show normal acquisition and retention of auditory fear conditioning (WT, open symbols). Labels indicate whether freezing was to tone (squares) or during the 2 min before tone (circles). All results are means $\pm$ s.e.m.

in the basal state but are induced by repeated tetani. L-LTP in WT slices therefore requires a stronger stimulation and is inhibited only at later times. Thus, two mechanisms underlie L-LTP in $GCN2^{-/-}$ slices: first, translation of pre-existing transcripts immediately increases LTP, and second, increased transcription of specific mRNAs generates persistent L-LTP.

How does GCN2 affect synaptic plasticity and learning? One possible model is based on the translational regulation of ATF4 mRNA through the GCN2-mediated phosphorylation of eIF2 α . A pivotal point is that ATF4 represses neuronal CREB activity^{1,7,8,10}. Thus, under basal conditions, when GCN2 and eIF2 α are phosphorylated and ATF4 levels are high, CREB-dependent transcription, synaptic plasticity and learning are repressed. By decreasing the phosphorylation of GCN2 and eIF2 α , LTP-inducing stimulation would remove this inhibition of synaptic plasticity and memory formation. In this manner, GCN2 regulates the switch from short-term to long-term memory.

We documented a correlation between L-LTP and spatial memory. In accordance with the low threshold for L-LTP and its suppression after strong stimulation, the spatial memory of $GCN2^{-/-}$ mice depended on the intensity of training: it was impaired by strong training and enhanced by weaker training. A likely explanation is that strong stimulation (behavioural or by four trains in slices) potentiates an inhibitory pathway that is facilitated in $GCN2^{-/-}$ mice. The nature of this mechanism will be an important target of future studies and may involve changes in regulation of gene expression and/or synaptic translation. Our results indicate that neurons might have not only a threshold for activation of gene expression but also a second threshold at which too much gene expression blocks synaptic

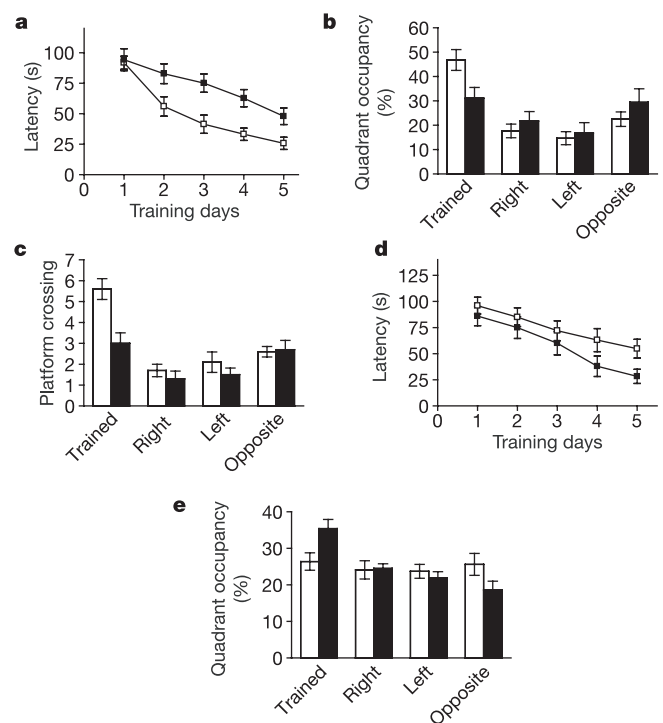


Figure 4 | Long-term spatial memory of $GCN2^{-/-}$ mice is enhanced after weak training but impaired after more intense training (in the Morris water maze). **a**, Escape latencies in hidden-platform tests (three trials a day), plotted as a function of training days (open squares, WT, $n = 16$; filled squares, $GCN2^{-/-}$, $n = 15$), are shorter for WT than $GCN2^{-/-}$ mice. **b**, After completion of training, WT mice (open bars) showed preferential quadrant occupancy in comparison with $GCN2^{-/-}$ mice (filled bars). **c**, WT mice (open bars) crossed the previous site where the platform was located more times than $GCN2^{-/-}$ mice (filled bars; $P < 0.001$). **d**, When locating the hidden platform (one trial a day), escape latencies were consistently shorter for $GCN2^{-/-}$ mice (filled squares) than for WT mice (open squares) ($n = 15$ for both). **e**, In the occupancy test, $GCN2^{-/-}$ mice (filled bars) spent more time in the trained quadrant than WT mice (open bars). All results are means $\pm$ s.e.m.

plasticity. Shutting off plasticity could be important under conditions of excessive activity such as seizures. Our results provide genetic evidence that translational control by GCN2 is critical for synaptic plasticity, learning and memory. In addition, they raise the prospect that memory formation is regulated through the translational control of transcription.

METHODS

Generation of transgenic mice by GCN2.KO4 targeting. *GCN2* was deleted by a targeting vector constructed from polymerase chain reaction (PCR) fragments amplified from cloned 129SvEv genomic DNA (see Supplementary Methods). Chimaeric mice derived from *GCN2.KO4*ex/+ embryonic stem cells were prepared by blastocyst injection and the mutant allele was transmitted through the germline to isogenic 129SvEv mice, which were bred to homozygosity. $GCN2^{-/-}$ mice were phenotypically normal in comparison with their wild-type littermates and were obtained in a mendelian ratio.

In situ hybridization histochemistry. Mouse sense and antisense cRNA probes coding for exon 12 of *GCN2* were labelled with [³⁵S]UTP and [³⁵S]CTP (1,250 Ci mmol⁻¹; Amersham), and *in situ* hybridization histochemistry was performed as reported previously²⁷.

Immunoprecipitation, immunohistochemistry and western blotting. Antibodies against the carboxy-terminal portion (kinase domain) of mouse *GCN2* kinase (C-term) and ATF4 have been described⁶. The antibody against the amino-terminal portion (amino acid residues 1–363; N-term) of human *GCN2* was produced as a glutathione *S*-transferase fusion protein in BL-21, and purified on glutathione-Sepharose (APB). Immunoblotting and immunohistochemistry were as reported^{6,28}. Antibodies against phospho-eIF2 α total eIF2 α and β -actin were purchased from Cell Signalling and Technology Laboratories.

Electrophysiology. After decapitation of WT ($GCN2^{+/+}$) or transgenic ($GCN2^{-/-}$) age-matched littermates (6–12 weeks old), hippocampal slices 400 μm thick were cut with a vibratome and kept submerged at 27–28 °C. Slices were perfused (at 1–2 ml min^{-1}) with oxygenated (95% O_2 , 5% CO_2) artificial cerebrospinal fluid (ACSF) containing 124 mM NaCl, 2.5 mM KCl, 1.25 mM NaH_2PO_4 , 1.3 mM MgSO_4 , 2.5 mM CaCl_2 , 26 mM NaHCO_3 and 10 mM glucose. Bipolar tungsten electrodes were placed in CA1 stratum radiatum to stimulate Schaffer collateral and commissural fibres, and extracellular field EPSPs (fEPSPs) were recorded from stratum radiatum with a glass microelectrode (2–3 M Ω , filled with 2 M NaCl). Stimulus (0.1-ms duration) was adjusted to evoke 35–40% maximal fEPSPs at 0.033 Hz. LTP was induced with one or four trains (1 s) at 100 Hz delivered 5 min apart. For LTD experiments, 1 Hz stimulation was applied for 15 min. Forskolin (50 μM ; Sigma) or DHPG (50 μM ; Tocris) was added to the bath after at least 30 min of stable recording. Anisomycin (40 μM ; Calbiochem), actinomycin D (40 μM ; Calbiochem) or KT5720 (1 μM ; Biomol) was applied for 30 min, or as indicated otherwise, before tetanic stimulation. Statistical analysis used *t*-tests and two-way ANOVA. All data are presented as means $\pm$ s.e.m.; *n* indicates the number of slices. The experimenter was blind to the mouse genotype.

Fear conditioning. The experimenter (blind to mouse genotype) compared $GCN2^{-/-}$ and WT littermates (males, 2–4 months old). Training consisted of two pairings of a tone (2,800 Hz, 85 dB, 30 s) with a co-terminating footshock (0.7 mA, 2 s). The first tone started 120 s after animals had been placed in the conditioning chamber, where they remained for a further 1 min after the second pairing, and were then returned to their home cage. Mice were tested 1 and 10 days later for freezing in response to training context in a counterbalanced manner (Supplementary Information).

Morris water maze task. The pool was 100 cm in diameter and the water was rendered opaque by the addition of white tempera. Water temperature was kept at 20 °C. The platform was 4.5 cm in diameter.

Mice were trained three times a day at intervals of 30 min, or once a day over five consecutive days. In each trial the mouse swam until it found the platform, or after 120 s it was guided to the platform; the mouse was then placed on the platform for 10 s before being picked up. At the end of the testing period, a probe trial (60 s) was performed. Statistical analysis was based on univariate and multivariate ANOVA, and between-group comparisons were made by Tukey's test.

Polysome profile analysis and RT–PCR. Hippocampal slices were washed twice with cold PBS containing 100 $\mu\text{g ml}^{-1}$ cycloheximide, suspended in lysis buffer, homogenized with 15 strokes (7-ml Wheaton Dounce) on ice, and then centrifuged for 2 min at 14,000g. Gradients were prepared and analysed as described²⁹. For detection of ATF4 and β -actin mRNAs, RNA from individual fractions was amplified in one-tube RT–PCR reactions, which were optimized to detect the exponential phase on the amplification curve.

Quantitative RT–PCR. The one-step RT–PCR LightCycler RNA Master SYBR Green kit (Roche) was used to quantify CRE-dependent gene expression, as recommended by the manufacturer. Primers, RT–PCR conditions and normalization procedures were as described³⁰.

Received 24 April; accepted 10 June 2005.

- Kandel, E. R. The molecular biology of memory storage: a dialogue between genes and synapses. *Science* **294**, 1030–1038 (2001).
- McGaugh, J. L. Memory—a century of consolidation. *Science* **287**, 248–251 (2000).
- Mathews, M. B., Sonenberg, N. & Hershey, J. W. B. in *Translational Control of Gene Expression* (eds Sonenberg, N., Hershey, J. W. B. & Mathews, M. B.) 1–33 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2000).
- Hinnebusch, A. G. in *Translational Control of Gene Expression* (eds Sonenberg, N., Hershey, J. W. B. & Mathews, M. B.) 185–244 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2000).
- Sonenberg, N. & Dever, T. E. Eukaryotic translation initiation factors and regulators. *Curr. Opin. Struct. Biol.* **13**, 56–63 (2003).
- Harding, H. P. *et al.* Regulated translation initiation controls stress-induced gene expression in mammalian cells. *Mol. Cell* **6**, 1099–1108 (2000).
- Yin, J. C. *et al.* Induction of a dominant negative CREB transgene specifically blocks long-term memory in *Drosophila*. *Cell* **79**, 49–58 (1994).
- Bartsch, D. *et al.* Aplysia CREB2 represses long-term facilitation: relief of repression converts transient facilitation into long-term functional and structural change. *Cell* **83**, 979–992 (1995).
- Abel, T., Martin, K. C., Bartsch, D. & Kandel, E. R. Memory suppressor genes: inhibitory constraints on the storage of long-term memory. *Science* **279**, 338–341 (1998).
- Chen, A. *et al.* Inducible enhancement of memory storage and synaptic plasticity in transgenic mice expressing an inhibitor of ATF4 (CREB-2) and C/EBP proteins. *Neuron* **39**, 655–669 (2003).
- Vattem, K. M. & Wek, R. C. Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. *Proc. Natl Acad. Sci. USA* **101**, 11269–11274 (2004).
- Scheuner, D. *et al.* Translational control is required for the unfolded protein response and *in vivo* glucose homeostasis. *Mol. Cell* **7**, 1165–1176 (2001).
- Santoyo, J., Alcalde, J., Mendez, R., Pulido, D. & de Haro, C. Cloning and characterization of a cDNA encoding a protein synthesis initiation factor-2 α (elf-2 α) kinase from *Drosophila melanogaster*. Homology to yeast GCN2 protein kinase. *J. Biol. Chem.* **272**, 12544–12550 (1997).
- Berlanga, J. J., Santoyo, J. & De Haro, C. Characterization of a mammalian homolog of the GCN2 eukaryotic initiation factor 2 α kinase. *Eur. J. Biochem.* **265**, 754–762 (1999).
- Sood, R., Porter, A. C., Olsen, D. A., Cavener, D. R. & Wek, R. C. A mammalian homologue of GCN2 protein kinase important for translational control by phosphorylation of eukaryotic initiation factor-2 α . *Genetics* **154**, 787–801 (2000).
- Frey, U. & Morris, R. G. Synaptic tagging and long-term potentiation. *Nature* **385**, 533–536 (1997).
- Kelleher, R. J. III, Govindarajan, A., Jung, H. Y., Kang, H. & Tonegawa, S. Translational control by MAPK signalling in long-term synaptic plasticity and memory. *Cell* **116**, 467–479 (2004).
- Huang, Y. Y. & Kandel, E. R. D1/D5 receptor agonists induce a protein synthesis-dependent late potentiation in the CA1 region of the hippocampus. *Proc. Natl Acad. Sci. USA* **92**, 2446–2450 (1995).
- Palmer, M. J., Irving, A. J., Seabrook, G. R., Jane, D. E. & Collingridge, G. L. The group I mGlu receptor agonist DHPG induces a novel form of LTD in the CA1 region of the hippocampus. *Neuropharmacology* **36**, 1517–1532 (1997).
- Impey, S. *et al.* Induction of CRE-mediated gene expression by stimuli that generate long-lasting LTP in area CA1 of the hippocampus. *Neuron* **16**, 973–982 (1996).
- LeDoux, J. E. Emotion circuits in the brain. *Annu. Rev. Neurosci.* **23**, 155–184 (2000).
- Morris, R. G., Garrud, P., Rawlins, J. N. & O'Keefe, J. Place navigation impaired in rats with hippocampal lesions. *Nature* **297**, 681–683 (1982).
- Barco, A., Alarcon, J. M. & Kandel, E. R. Expression of constitutively active CREB protein facilitates the late phase of long-term potentiation by enhancing synaptic capture. *Cell* **108**, 689–703 (2002).
- Malleret, G. *et al.* Inducible and reversible enhancement of learning, memory, and long-term potentiation by genetic inhibition of calcineurin. *Cell* **104**, 675–686 (2001).
- Chen, C. Y. & Shyu, A. B. Selective degradation of early-response-gene mRNAs: functional analyses of sequence features of the AU-rich elements. *Mol. Cell. Biol.* **14**, 8471–8482 (1994).
- Shyu, A. B., Greenberg, M. E. & Belasco, J. G. The c-fos transcript is targeted for rapid decay by two distinct mRNA degradation pathways. *Genes Dev.* **3**, 60–72 (1989).
- Seidah, N. G. *et al.* The secretory proprotein convertase neural apoptosis-regulated convertase 1 (NARC-1): liver regeneration and neuronal differentiation. *Proc. Natl Acad. Sci. USA* **100**, 928–933 (2003).
- Lapointe, V. *et al.* Synapse-specific mGluR1-dependent long-term potentiation in interneurons regulates mouse hippocampal inhibition. *J. Physiol. (Lond.)* **555**, 125–135 (2004).
- Chan, J., Khan, S. N., Harvey, I., Merrick, W. & Pelletier, J. Eukaryotic protein synthesis inhibitors identified by comparison of cytotoxicity profiles. *RNA* **10**, 528–543 (2004).
- Saura, C. A. *et al.* Loss of presenilin function causes impairments of memory and synaptic plasticity followed by age-dependent neurodegeneration. *Neuron* **42**, 23–36 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank E. Kandel, K. Krnjević, K. Rosenblum, E. Landau, R. Blitzer, C. Alberini, Y. Mamane and T. Lubell for comments on the manuscript; Y. Zhang, R. Jungreis and A. Sylvestre for assisting in the production and maintenance of the $GCN2^{-/-}$ mice; and Colin Lister for assistance. This work was supported by grants from the Canadian Institute of Health Research (CIHR) and the Howard Hughes Medical Institute (HHMI) to N.S., a CIHR Group Grant to J.-C.L. and W.S.; a CIHR grant to N. Seidah; an NIH grant to D.R.; CIHR, Natural Sciences and Engineering Research Council of Canada (NSERC), Volkswagen Foundation and EJLB Foundation grants to K.N.; and a CIHR grant to A.C.C. N.S. is a CIHR Distinguished Scientist and a HHMI International scholar. M.C.-M. is supported by a CIHR postdoctoral fellowship.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to N.S. (nahum.sonenberg@mcgill.ca).

Aminoglycoside antibiotics induce bacterial biofilm formation

Lucas R. Hoffman¹, David A. D'Argenio⁴, Michael J. MacCoss², Zhaoying Zhang⁵, Roger A. Jones⁵ & Samuel I. Miller^{2,3,4}

Biofilms are adherent aggregates of bacterial cells that form on biotic and abiotic surfaces, including human tissues. Biofilms resist antibiotic treatment and contribute to bacterial persistence in chronic infections^{1,2}. Hence, the elucidation of the mechanisms by which biofilms are formed may assist in the treatment of chronic infections, such as *Pseudomonas aeruginosa* in the airways of patients with cystic fibrosis². Here we show that subinhibitory concentrations of aminoglycoside antibiotics induce biofilm formation in *P. aeruginosa* and *Escherichia coli*. In *P. aeruginosa*, a gene, which we designated aminoglycoside response regulator (*arr*), was essential for this induction and contributed to biofilm-specific aminoglycoside resistance. The *arr* gene is predicted to encode an inner-membrane phosphodiesterase whose substrate is cyclic di-guanosine monophosphate (c-di-GMP)—a bacterial second messenger that regulates cell surface adhesiveness³. We found that membranes from *arr* mutants had diminished c-di-GMP phosphodiesterase activity, and *P. aeruginosa* cells with a mutation changing a predicted catalytic residue of Arr were defective in their biofilm response to tobramycin. Furthermore, tobramycin-inducible biofilm formation was inhibited by exogenous GTP, which is known to inhibit c-di-GMP phosphodiesterase activity⁴. Our results demonstrate that biofilm formation can be a specific, defensive reaction to the presence of antibiotics, and indicate that the molecular basis of this response includes alterations in the level of c-di-GMP.

Most antibiotics of clinical relevance are derivatives of naturally occurring microbial products that probably function in microbial competition within environmental niches⁵. The aminoglycosides are a class of clinically important antibiotics that have been widely used to treat chronic bacterial infections of the heart, lung and urinary tract⁶. Tobramycin, an aminoglycoside produced by the bacterium *Streptomyces tenebrarius*, is commonly used because of its enhanced effectiveness against infections with the opportunistic pathogen *P. aeruginosa* (ref. 6). Nevertheless, clinical isolates of *P. aeruginosa* possess an inducible resistance to tobramycin⁷, suggesting that *P. aeruginosa* responds adaptively to aminoglycoside exposure. Because both *P. aeruginosa* and *S. tenebrarius* are present in soil⁸, we proposed that *P. aeruginosa* had evolved adaptive responses to tobramycin before the clinical use of antibiotics, and that one such response was the formation of antibiotic-resistant biofilms. This response would be advantageous for *P. aeruginosa* both in the soil, when encountering an aminoglycoside-producing bacterium such as *S. tenebrarius*, and in human hosts receiving antibiotic therapy, where bacteria may encounter variable antibiotic concentrations⁹. Such situations include the chronic airway infections of cystic fibrosis (CF) patients, in which *P. aeruginosa* frequently reaches densities of 10⁹ viable cells per ml of sputum¹⁰. Upon treatment with tobramycin in aerosol form, a fraction of these bacteria may inevitably be

exposed to subinhibitory levels of antibiotic. Indeed, tobramycin rarely eradicates these infections, even before the appearance of tobramycin-resistant mutants¹⁰.

To study the earliest physiological response of *P. aeruginosa* to tobramycin, we used concentrations ≤ 0.3 times the minimal inhibitory concentration (MIC) of 1 $\mu\text{g ml}^{-1}$. These tobramycin concentrations did not measurably alter the growth rate of *P. aeruginosa* strain PAO1 in Mueller-Hinton broth (MHB; Fig. 1a and Supplementary Information); nor did they alter global protein synthesis as determined by measurement of total cellular protein in the PAO1 strain and β -galactosidase activity in strains with *lacZ* translational fusions to several genes (*hcnB*, *phzD*, *pqsH*, *katB*; data not shown). Despite this lack of inhibition, a range of subinhibitory tobramycin concentrations induced biofilm formation in PAO1 (Fig. 1b, c). A peak induction in biofilm mass of 3.4-fold occurred during growth in

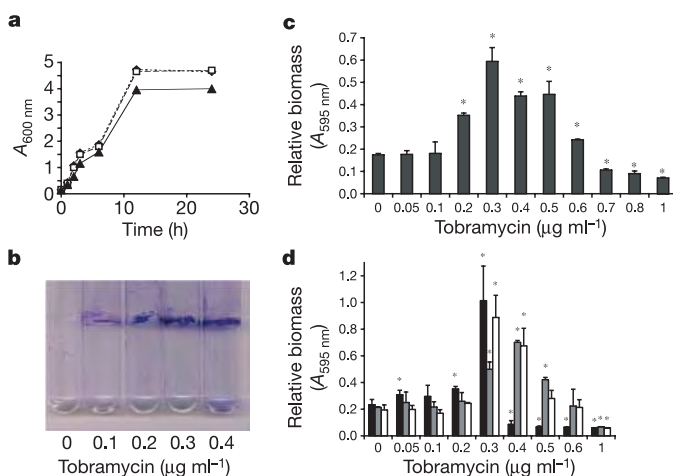


Figure 1 | Phenotypic effects of tobramycin on *P. aeruginosa* and *E. coli*. **a**, Growth curve of shaken cultures of *P. aeruginosa* strain PAO1 in 0 (filled diamonds), 0.3 (open squares) and 0.4 (filled triangles) $\mu\text{g ml}^{-1}$ tobramycin (MIC 1 $\mu\text{g ml}^{-1}$). **b**, Tobramycin-induced biofilm formation in PAO1 cultured on glass culture tubes (after 24 h static growth) visualized by crystal-violet staining. **c**, Tobramycin-induced biofilm formation in PAO1 cultured on plastic microtitre plates. Note that biofilm formation in the absence of the drug is lower on glass (panel **b**) than in polystyrene plates. Results are averages of 8 replicates $\pm$ s.e.m. and are representative of 10 independent experiments. Asterisk, $P < 0.001$ as compared with PAO1 exposed to no drug. **d**, Tobramycin-induced biofilm formation in *E. coli* cultured on plastic microtitre plates. Three strains were tested, with the results for individual strains indicated by black, grey or white bars, respectively. Results are averages of 8 replicates $\pm$ s.e.m. Asterisk, $P < 0.001$ as compared with the same strain exposed to no drug.

¹Department of Pediatrics, ²Department of Genome Sciences, ³Department of Medicine and ⁴Department of Microbiology, University of Washington, Seattle, Washington 98195, USA. ⁵Department of Chemistry and Department of Chemical Biology, Rutgers University, Piscataway, New Jersey 08854, USA.

the presence of tobramycin at 0.3 times the MIC (Fig. 1c)—an increase in biomass similar to that associated with *P. aeruginosa* mutants with increased propensities for biofilm formation^{2,11}. Tobramycin induced biofilm formation in three different growth media (M63 salts supplemented with glucose and casamino acids, MHB, Luria broth), on several abiotic surfaces (borosilicate glass, polystyrene, polypropylene, polycarbonate) and in 13 out of 14 *P. aeruginosa* clinical isolates tested (Fig. 1b, c and Supplementary Information and data not shown).

For both Gram-negative and Gram-positive bacteria, subinhibitory antibiotic treatment can stimulate production of exopolysaccharides^{12,13}. Subinhibitory levels of the β -lactam antibiotic imipenem augment production of the *P. aeruginosa* exopolysaccharide alginate, and leads to increased biofilm volume¹³. However, *P. aeruginosa* *algD* mutants (which are defective in alginate production) and the parental strain PAO1 demonstrated equivalent tobramycin-induced biofilm formation (data not shown). Furthermore, the tobramycin-induced increase in biomass (Fig. 1c) corresponded to an increase in biofilm colony-forming units (CFU) (Supplementary Information), indicating that it resulted primarily from an increase in cell number and not an increase in extracellular matrix.

We assayed other antibiotics at a range of concentrations (relative to the MIC) equivalent to those used for tobramycin. Three additional aminoglycoside antibiotics induced biofilm formation, although tobramycin had the strongest effect. The maximum induction by amikacin, streptomycin and gentamicin was 75%, 66% and 25% of that by tobramycin, respectively. In contrast, polymyxin B (a peptide antibiotic that is cationic like aminoglycosides and interacts with membranes) had no effect on biofilm formation; nor did the protein synthesis inhibitor chloramphenicol or the cell wall synthesis inhibitor carbenicillin (data not shown). Therefore, biofilm induction by tobramycin is unlikely to be solely due to non-specific protein synthesis inhibition, cell damage or interaction with cell membranes through positive charge; rather, it appears to be a specific response by *P. aeruginosa* to aminoglycosides.

In order to test whether this response was present in other Gram-negative bacteria, we examined the enteric bacterium *E. coli*. Subinhibitory levels of tobramycin induced biofilm formation in three clinical isolates of *E. coli* from three patients with bacteraemia (Fig. 1d). As with *P. aeruginosa*, this induction occurred in a range of tobramycin concentrations centred on $0.3 \mu\text{g ml}^{-1}$ ($0.3 \times$ the MIC for *E. coli*). Such conservation suggests that tobramycin activates a signalling pathway present in both *P. aeruginosa* and *E. coli*.

Signalling by the second messenger c-di-GMP is a good candidate for such a conserved system as this dinucleotide regulates cell adhesiveness in a diverse range of bacteria^{3,14,15}. The PAO1 genome includes at least 38 genes that are predicted to encode a regulator of intracellular c-di-GMP levels. These proteins contain either a GGDEF domain (found in putative cyclases for c-di-GMP synthesis), an EAL domain (found in putative phosphodiesterases for c-di-GMP degradation), or both¹⁶. To identify which, if any, of these genes might be involved in signalling the presence of tobramycin, we screened the relevant transposon-insertion mutants of PAO1 (ref. 17) for tobramycin-inducible biofilm formation. We reasoned that the inactivation of a gene involved in signalling the presence of tobramycin would result in a strain that showed reduced biofilm induction by tobramycin, but normal biofilm production in the absence of tobramycin, relative to wild-type cells. Many of the mutants had altered biofilm formation even in the absence of tobramycin (Supplementary Information), and these were not analysed further.

Three strains, each with a different insertion mutation in the monocistronic open reading frame PA2818, were defective for tobramycin-induced biofilm formation (Fig. 2a, b and Supplementary Information). We designated this gene *arr*, for aminoglycoside response regulator. The *arr* gene is predicted to encode an inner-

membrane protein with two transmembrane domains, a periplasmic domain that could transduce an environmental stimulus, and an EAL domain (Fig. 3a). Database searches revealed a similar domain architecture in the gene products of many Gram-negative bacteria, including *E. coli*. Complementation with a plasmid expressing *arr* restored wild-type tobramycin-induced biofilm formation in an *arr* mutant (Fig. 2a), indicating that Arr is necessary for this response. There was no complementation with a plasmid expressing a mutant *arr* gene encoding a protein in which the conserved glutamate residue of the EAL domain was replaced with an alanine (E297A) (Fig. 2d). As this mutation has abolished the biological activity attributed to other EAL domain proteins¹⁸, this result suggested that Arr phosphodiesterase activity is required for tobramycin-induced biofilm formation.

Consistent with Arr having c-di-GMP phosphodiesterase activity, membranes from *arr* mutant cells were 54% less active in degrading c-di-GMP than membranes from PAO1 cells (Fig. 3b); this is a surprising finding given that 23 of the 38 predicted PAO1 c-di-GMP regulators are expected to be transmembrane proteins. Wild-type phosphodiesterase activity was restored to the *arr* mutant by a plasmid expressing *arr* (Fig. 3b), but not by a plasmid expressing the gene encoding the E297A Arr mutant (data not shown). Furthermore, tobramycin-inducible biofilm formation in wild-type cells was inhibited by exogenous GTP—a c-di-GMP phosphodiesterase inhibitor⁴ (Supplementary Information). Based on these results, we propose that tobramycin, either directly or indirectly, enhances the phosphodiesterase activity of the Arr cytoplasmic EAL domain, leading to c-di-GMP inactivation and augmented biofilm formation (Fig. 3a).

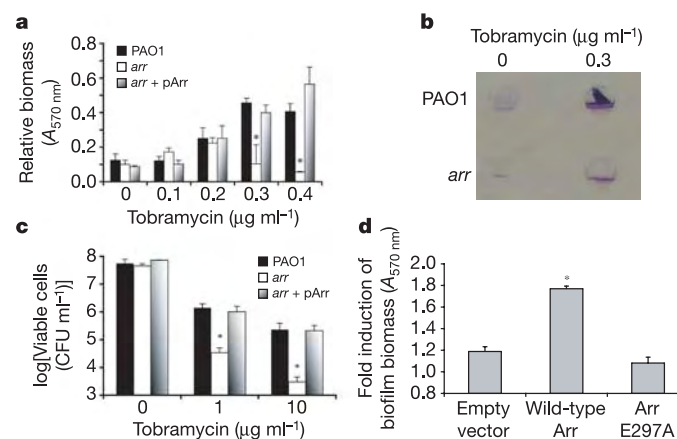


Figure 2 | The role of *P. aeruginosa* *arr* in tobramycin induction of biofilm formation and biofilm antibiotic resistance. **a**, Microtitre plate biofilm assay with the parental PAO1 strain, the *arr* mutant strain, and the *arr* mutant strain carrying a plasmid expressing the wild-type *arr* gene (pArr). Results are averages of 8 replicates $\pm$ s.e.m. and are representative of three independent experiments. Asterisk, $P < 0.001$ as compared with PAO1 cultured in the same tobramycin concentration. **b**, Crystal-violet-stained biofilms of PAO1 and the *arr* mutant strain grown for 24 h on glass coverslips spanning the air–liquid interface of standing cultures. The dark portion above the air–liquid interface biofilm in tobramycin-induced PAO1 is adherent pellicle. **c**, Microtitre plate biofilm antibiotic resistance assay in which biofilms of the indicated strains were challenged with tobramycin at the concentrations shown, and the survival of suspended biofilm cells determined. Results are the average of three experiments $\pm$ s.e.m. Asterisk, $P < 0.001$ as compared with PAO1 biofilms exposed to the same drug concentration. **d**, The indicated plasmids were introduced into the *arr* mutant strain. Fold induction of biofilm formation by $0.3 \mu\text{g ml}^{-1}$ tobramycin for each transformant was determined (average $\pm$ s.e.m. of 6 replicates). Asterisk, $P < 0.001$ as compared with cells carrying the empty vector. Cells were grown in $500 \mu\text{g ml}^{-1}$ carbenicillin for plasmid maintenance.

Although Arr promotes biofilm formation, expression of EAL-type regulators in other organisms has been linked to a reduction in biofilm formation¹⁵; this is in apparent contradiction with our model (Fig. 3a). However, inactivation of putative c-di-GMP regulatory proteins (with either GGDEF, EAL, or both domains) had remarkably varied effects on basal levels of biofilm formation in PAO1 (Supplementary Information). This indicates a complex relationship between biofilm formation and the expression of individual regulators, which could involve the localized regulation of discrete cytoplasmic c-di-GMP pools. The *Caulobacter crescentus* GGDEF-type regulator PleD (ref. 14) and the *P. aeruginosa* GGDEF/EAL-type regulator FimX are both localized to one cell pole^{3,14,19}. Similarly, the Arr periplasmic domain could mediate the formation of localized protein complexes.

Multiple cell surface appendages mediate bacterial aggregation and facilitate biofilm formation. Flagella and pili are involved in biofilm formation^{11,20} and are regulatory targets of c-di-GMP (refs 14, 19). It was possible that tobramycin and Arr were regulating biofilm formation by altering these surface appendages. Indeed, subinhibitory tobramycin concentrations inhibited PAO1 flagellar motility in a dose-dependent manner, and inhibition was reduced in an *arr* mutant (Supplementary Information). Nevertheless, tobramycin did not affect type IV pili-mediated twitching motility, and PAO1 *fliC* mutants (lacking flagella) and *pilA* mutants (lacking type IV pili) showed tobramycin-induced biofilm formation that was equivalent

to wild-type levels (data not shown). The *arr* mutant was also motile in assays for twitching (data not shown) and flagellar swimming (Supplementary Information), and polymyxin B inhibited flagellar swimming (Supplementary Information) without any effect on biofilm formation (data not shown). Therefore, it is unlikely that tobramycin or Arr affect biofilm formation simply via these surface appendages.

Because biofilms are associated with increased antibiotic resistance^{1,2}, we tested whether Arr had a role in biofilm-mediated antibiotic resistance as well as biofilm formation. In a peg biofilm assay that was recently proposed for clinical use²¹, biofilms of *arr*-mutant cells were approximately 100-fold more susceptible to tobramycin killing than PAO1 biofilms, and the wild-type phenotype was restored by expressing *arr* from a plasmid (Fig. 2c). Decreased biofilm tobramycin-resistance of the *arr* mutant was also observed in a standard biofilm crystal violet staining assay (data not shown), and by measuring CFU in colony biofilms (Supplementary Information). In contrast, planktonic cultures of *arr*-mutant cells exhibited the same killing by tobramycin as cultures of wild-type cells (MIC of $1 \mu\text{g ml}^{-1}$). This indicates that *arr*, like *ndvB* (ref. 1), is a genetic determinant of biofilm-mediated antibiotic resistance in *P. aeruginosa*. The contribution of Arr to biofilm tobramycin-resistance could be clinically relevant, as improvement of lung function in CF patients treated with tobramycin in aerosol form correlates with an approximate 100-fold reduction in *P. aeruginosa* CFU (ref. 10); this is equivalent to the difference in killing between wild-type and *arr*-mutant biofilms treated with tobramycin (Fig. 2c).

Another *P. aeruginosa* EAL-type regulator, PvrR, plays a role in the formation of small colony variants (SCVs) that are hyperadherent and antibiotic resistant²². Mutation of *pvrR* increased the frequency of appearance of SCVs growing in the presence of high concentrations of the aminoglycoside kanamycin²². This raised the possibility that mutation of *arr* altered the frequency of such variants with a consequent change in tobramycin-induced biofilm formation. To test this, we isolated and examined resuspended cells from PAO1 and *arr*-mutant biofilms grown with and without subinhibitory tobramycin. The frequency of SCVs was equivalent in each of the four cell populations (data not shown). In addition, we tested these resuspended cells for their ability to form new biofilms in the absence of antibiotic, and for their tobramycin susceptibilities in both static and shaken planktonic cultures. None of these characteristics were altered, either by mutation of *arr* or by subinhibitory tobramycin treatment (Supplementary Information). Therefore, subinhibitory tobramycin does not enrich for relatively resistant or adherent genetic variants, and the *arr* mutant phenotype is not due to the altered frequency of such variants.

Taken together, our results suggest that inhibiting the activity of EAL-type regulators such as Arr might be of therapeutic benefit early in *P. aeruginosa* chronic infections: particularly airway infections in which the aerosol form of tobramycin is widely used. It might also be beneficial in acute disease, as the mutation of *pvrR* decreased *P. aeruginosa* virulence in burned mice²³. However, clinical strain variability could complicate such therapeutic manipulation of c-di-GMP metabolism. The *pvrR* locus of strain PA14 is contained on a genetic island that is absent in strain PAO1 (ref. 23). Similarly, a genomic microarray analysis suggested that *arr* was absent or divergent in some *P. aeruginosa* isolates²⁴. This variability notwithstanding, 13 of the 14 CF isolates we tested demonstrated tobramycin-induced biofilm formation, while the one remaining isolate appeared to lack *arr* by polymerase chain reaction (PCR) analysis (data not shown). Therefore, the relevant characteristic conserved among these strains and with *E. coli* (Fig. 1c, d) is antibiotic induction of biofilm formation, probably through c-di-GMP as a second messenger. This biofilm response could contribute to differences in therapeutic outcome upon antibiotic treatment, such as in CF where not all patients respond positively to tobramycin aerosols¹⁰, and the ability to manipulate c-di-GMP metabolism may have

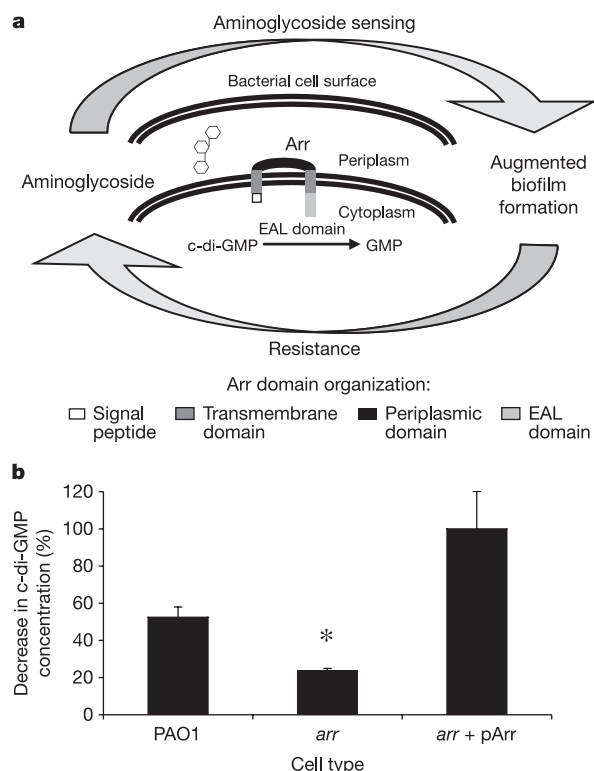


Figure 3 | The role of c-di-GMP in the biofilm response to aminoglycosides. **a**, A model for aminoglycoside effects on *P. aeruginosa*. Aminoglycoside antibiotics such as tobramycin (indicated by the three-ringed structure) could act as first messengers that trigger changes—mediated either by binding directly to proteins such as Arr or indirectly through intermediary molecules—in the level of the second messenger c-di-GMP. The proposed domain organization of Arr is indicated, including the location of the carboxy-terminal EAL domain. **b**, Effect on c-di-GMP concentration of washed cell membranes containing 0.4 mg of protein from either the PAO1 strain, the *arr* mutant strain, or the *arr* mutant strain with a plasmid expressing the wild-type *arr* gene after incubation with $10 \mu\text{M}$ c-di-GMP. Results are averages of at least 3 replicates $\pm$ s.d., and are representative of four separate experiments. Asterisk, $P < 0.001$ as compared with PAO1.

therapeutic value by increasing bacterial susceptibility to standard therapy.

There is growing evidence that bacteria respond specifically and defensively to subinhibitory antibiotic concentrations^{5,25}. The evidence presented here indicates that Gram-negative bacteria can respond to aminoglycosides by forming antibiotic-resistant biofilms—perhaps only one of many strategies used to counter antibiotic production by Gram-positive soil bacteria such as the Streptomyces. In *P. aeruginosa*, this biofilm response requires Arr—a regulator that alters c-di-GMP levels (Fig. 3a). Both in soil and within animals, c-di-GMP regulation could contribute to the diversity of possible outcomes for bacterial communities challenged with antibiotics.

METHODS

Bacterial strains, chemicals and media. The PAO1 strain was obtained from B. Iglewski through C. Manoil, and the transposon-insertion mutants from M. Jacobs¹⁷. *E. coli* isolates were from patients with bacteraemia at the Massachusetts General Hospital (archived by S. Miller), and *P. aeruginosa* longitudinal clinical isolates were from young CF patients less than 8 yr old (provided by J. Burns²⁶). PAO1 mutants containing a transposon TN5-derived insertion element¹⁷ in the gene PA2818 that were analysed in detail included the 50022 (PA2818::ISphoA/hah-Tc), 17026 (PA2818::ISlacZ/hah-Tc) and 5339 (PA2818::ISlacZ/hah-Tc) mutants; these mutants are described more fully at <http://www.genome.washington.edu/UWGC/pseudomonas/index.cfm>. Mutants in other genes are described in the Supplementary Methods. All data shown for mutants in the PA2818 gene are for PAO1 mutant 50022; results were confirmed with PAO1 mutants 17026 and 5339. Bacteria were grown at 37 °C in MHB (Difco) unless otherwise indicated. Polymyxin B was purchased from Amersham Life Sciences. GTP, as well as tobramycin sulphate and all other antibiotics, were obtained from Sigma. Unlabelled and ¹⁸O-labelled c-di-GMP was synthesized, purified and characterized as described²⁷.

Phenotypic assays. *P. aeruginosa* PAO1 was inoculated in duplicate onto Mueller-Hinton agar (MHA) with 0.3% agar (to characterize swimming motility) or beneath MHA with 1.5% agar (to characterize twitching motility), each with and without antibiotics, and then incubated at 37 °C for 12 h, essentially as described²⁰. For use as an inoculum for MIC determination (by broth microdilution), killing assays, growth curves and biofilm assays, overnight cultures of bacteria grown in MHB were diluted approximately 1:100 with MHB (to a density of 10⁷ CFU ml⁻¹). Antibiotics were added to this inoculum as indicated. Cell number was determined by measuring CFU (by plating serial dilutions of cultures onto LB agar) or by measuring the absorbance at 600 nm (A_{600 nm}) of suspended cells. Biofilm formation was routinely quantified by measuring either A_{570 nm} or A_{595 nm} (which gave equivalent results) of crystal-violet staining of adherent cells as described previously²¹. To assay the induction of biofilm formation, biofilms were grown for 24 h without shaking at 37 °C either in glass culture tubes, on glass cover slips, or in wells of untreated 96-well polystyrene microtitre plates (Nunc) using 100 µl of culture per well (8 duplicates per condition per experiment). To assay biofilm antibiotic resistance and to confirm biofilm induction, biofilms were grown in microtitre plates without antibiotics as above and with a lid containing 96 polystyrene pegs (Nunc) such that a peg was inserted into each well. The pegs with adherent biofilms were removed after 24 h, washed three times in water, and placed into a new microtitre plate with fresh MHB supplemented with antibiotics. After static incubation for 24 h at 37 °C, the pegs were removed, washed three times with water, inserted into a 96-well plate containing fresh MHB, centrifuged at 1,811 g for 20 min, and then sonicated at approximately 25 °C for 5 min (using a Branson 1510 water bath; Branson) to remove and disperse adherent cells for CFU determination as described²¹. Consistent removal of the biofilm was confirmed by crystal-violet staining of the pegs. For microtitre biofilm assays, reported values are the mean of at least three replicates (with the error calculated as s.e.m.). Student's two-tailed *t*-test was used to establish the significance of differences between two means. In parallel with each biofilm experiment, an MIC was determined for the planktonic cells in the culture used as the inoculum. For all of the *P. aeruginosa* strains derived from PAO1, and for the *E. coli* strains tested, the MIC of tobramycin was consistently 1 µg ml⁻¹, including for resuspended biofilm-grown *P. aeruginosa*.

Cloning of *arr* and complementation experiments. A PCR fragment containing the gene PA2818 (nucleotides 3,170,500–3,174,000 in the *P. aeruginosa* PAO1 genome) was cloned into the *Eco*RI and *Hind*III sites of pUCP18. The resulting plasmid was introduced into PAO1 and the PA2818 (*arr*) 50022 mutant by electroporation and selection for carbenicillin resistance using standard

methods. These transformants were found to retain the plasmid for over 36 h of growth in the absence of selection. Therefore, biofilm experiments with the strain carrying the resulting plasmid with the wild-type *arr* gene were performed in the absence of carbenicillin, except as indicated.

Measurement of c-di-GMP. Washed cell membranes were prepared and assayed for c-di-GMP degrading activity essentially as described²⁸. Overnight cultures of *arr* mutant and wild-type PAO1 were diluted 1:100 in MHB and grown statically at 37 °C for 18 h. The cells were collected by centrifugation, the pellet resuspended in TME buffer (50 mM Tris-HCl pH 8, 0.9 mM EDTA, 10 mM MgCl₂) and the cells disrupted by sonication on ice. Unbroken cells were removed by centrifuging at 3,000 r.p.m. for 5 min. Membranes were separated from the supernatant by centrifuging at 14,000 r.p.m. for 30 min. The resulting pellets of washed membranes were resuspended in TME buffer, and protein concentration determined by a modified Lowry assay (BioRad). Aliquots of membrane corresponding to the indicated amount of protein were resuspended in 200 µl TME buffer with 10 µM c-di-GMP. The mixture was incubated for 10 min at 30 °C, and the reaction halted by boiling at 100 °C for 2 min. Membranes were removed by centrifugation at 14,000 r.p.m. for 30 min. The supernatant was analysed for c-di-GMP content as described in the Supplementary Methods.

Received 23 March; accepted 9 June 2005.

- Mah, T. F. *et al.* A genetic basis for *Pseudomonas aeruginosa* biofilm antibiotic resistance. *Nature* **426**, 306–310 (2003).
- Whiteley, M. *et al.* Gene expression in *Pseudomonas aeruginosa* biofilms. *Nature* **413**, 860–864 (2001).
- D'Argenio, D. A. & Miller, S. I. Cyclic di-GMP as a bacterial second messenger. *Microbiol.* **150**, 2497–2502 (2004).
- Ross, P. *et al.* The cyclic diguanylic acid regulatory system of cellulose synthesis in *Acetobacter xylinum*. Chemical synthesis and biological activity of cyclic nucleotide dimer, trimer, and phosphothioate derivatives. *J. Biol. Chem.* **265**, 18933–18943 (1990).
- Goh, E. B. *et al.* Transcriptional modulation of bacterial gene expression by subinhibitory concentrations of antibiotics. *Proc. Natl Acad. Sci. USA* **99**, 17025–17030 (2002).
- Goodman, L. S. & Gilman, A. *Goodman and Gilman's The Pharmacological Basis of Therapeutics* (Macmillan, New York, 1985).
- Barclay, M. L. *et al.* Adaptive resistance to tobramycin in *Pseudomonas aeruginosa* lung infection in cystic fibrosis. *J. Antimicrob. Chemother.* **37**, 1155–1164 (1996).
- Bergey, D. H. & Holt, J. G. *Bergey's Manual of Systematic Bacteriology* (Williams & Wilkins, Baltimore, 1984).
- Mukhopadhyay, S. *et al.* The quantitative distribution of nebulized antibiotic in the lung in cystic fibrosis. *Respir. Med.* **88**, 203–211 (1994).
- Ramsey, B. W. *et al.* Intermittent administration of inhaled tobramycin in patients with cystic fibrosis. Cystic Fibrosis Inhaled Tobramycin Study Group. *N. Engl. J. Med.* **340**, 23–30 (1999).
- Chiang, P. & Burrows, L. L. Biofilm formation by hyperpilated mutants of *Pseudomonas aeruginosa*. *J. Bacteriol.* **185**, 2374–2378 (2003).
- Rachid, S., Ohlsen, K., Witte, W., Hacker, J. & Ziebuhr, W. Effect of subinhibitory antibiotic concentrations on polysaccharide intercellular adhesion expression in biofilm-forming *Staphylococcus epidermidis*. *Antimicrob. Agents Chemother.* **44**, 3357–3363 (2000).
- Bagge, N. *et al.* *Pseudomonas aeruginosa* biofilms exposed to imipenem exhibit changes in global gene expression and beta-lactamase and alginate production. *Antimicrob. Agents Chemother.* **48**, 1175–1187 (2004).
- Paul, R. *et al.* Cell cycle-dependent dynamic localization of a bacterial response regulator with a novel di-guanylate cyclase output domain. *Genes Dev.* **18**, 715–727 (2004).
- Simm, R., Morr, M., Kader, A., Nimtz, M. & Romling, U. GGDEF and EAL domains inversely regulate cyclic di-GMP levels and transition from sessility to motility. *Mol. Microbiol.* **53**, 1123–1134 (2004).
- Galperin, M. Y., Nikolskaya, A. N. & Koonin, E. V. Novel domains of the prokaryotic two-component signal transduction systems. *FEMS Microbiol. Lett.* **203**, 11–21 (2001).
- Jacobs, M. A. *et al.* Comprehensive transposon mutant library of *Pseudomonas aeruginosa*. *Proc. Natl Acad. Sci. USA* **100**, 14339–14344 (2003).
- Bobrov, A. G., Kirillina, O. & Perry, R. D. The phosphodiesterase activity of the HmsP EAL domain is required for negative regulation of biofilm formation in *Yersinia pestis*. *FEMS Microbiol. Lett.* **247**, 123–130 (2005).
- Huang, B., Whitchurch, C. B. & Mattick, J. S. FimX, a multidomain protein connecting environmental signals to twitching motility in *Pseudomonas aeruginosa*. *J. Bacteriol.* **185**, 7068–7076 (2003).
- O'Toole, G. A. & Kolter, R. Flagellar and twitching motility are necessary for *Pseudomonas aeruginosa* biofilm development. *Mol. Microbiol.* **30**, 295–304 (1998).
- Moskowitz, S. M., Foster, J. M., Emerson, J. & Burns, J. L. Clinically feasible biofilm susceptibility assay for isolates of *Pseudomonas aeruginosa* from patients with cystic fibrosis. *J. Clin. Microbiol.* **42**, 1915–1922 (2004).
- Drenkard, E. & Ausubel, F. M. *Pseudomonas* biofilm formation and antibiotic

- resistance are linked to phenotypic variation. *Nature* **416**, 740–743 (2002).
23. He, J. *et al.* The broad host range pathogen *Pseudomonas aeruginosa* strain PA14 carries two pathogenicity islands harboring plant and animal virulence genes. *Proc. Natl Acad. Sci. USA* **101**, 2530–2535 (2004).
 24. Wolfgang, M. C. *et al.* Conservation of genome content and virulence determinants among clinical and environmental isolates of *Pseudomonas aeruginosa*. *Proc. Natl Acad. Sci. USA* **100**, 8484–8489 (2003).
 25. Bader, M. W. *et al.* Regulation of *Salmonella typhimurium* virulence gene expression by cationic antimicrobial peptides. *Mol. Microbiol.* **50**, 219–230 (2003).
 26. Rosenfeld, M. *et al.* Early pulmonary infection, inflammation, and clinical outcomes in infants with cystic fibrosis. *Pediatr. Pulmonol.* **32**, 356–366 (2001).
 27. Hisert, K. M. *et al.* A glutamate-alanine-leucine (EAL) domain protein of *Salmonella* controls bacterial survival in mice, antioxidant defence and killing of macrophages: role of cyclic diGMP. *Mol. Microbiol.* **56**, 1234–1245 (2005).
 28. Amikam, D. & Benziman, M. Cyclic diguanylic acid and cellulose synthesis in *Agrobacterium tumefaciens*. *J. Bacteriol.* **171**, 6649–6655 (1989).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. K. Ernst for assistance with preparing the manuscript and laboratory techniques; T. Guina and M. Wu for help in two-dimensional protein electrophoresis analyses; J. Foster and S. Moskowitz for instruction in biofilm growth and measurement; M. Olson and M. Jacobs for providing transposon-insertion mutants; M. Bader for assistance with the cloning and complementation of *arr*; and J. Burns, B. Ramsey and R. Gibson for discussions. This work was supported by the Cystic Fibrosis Foundation and the National Institutes of Health.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.I.M. (millersi@u.washington.edu).

LETTERS

A new family of RhoGEFs activates the Rop molecular switch in plants

Antje Berken¹, Christoph Thomas¹ & Alfred Wittinghofer¹

In plants, the small GTP-binding proteins called Rops work as signalling switches that control growth, development and plant responses to various environmental stimuli^{1–3}. Rop proteins (Rho of plants, Rac-like and AtRac in *Arabidopsis thaliana*) belong to the Rho family of Ras-related GTP-binding proteins that turn on signalling pathways by switching from a GDP-bound inactive to a GTP-bound active conformation^{4,5}. Activation depends on guanine nucleotide exchange factors (GEFs) that catalyse the otherwise slow GDP dissociation for subsequent GTP binding⁶. Although numerous RhoGEFs exist in animals and yeasts⁷, no Rop-specific GEFs have yet been identified in plants and so Rop activation has remained elusive^{1–3,8}. Here we describe a new family of RhoGEF proteins that are exclusive to plants. We define a unique domain within these RopGEFs, termed PRONE (plant-specific Rop nucleotide exchanger), which is exclusively active towards members of the Rop subfamily. It increases nucleotide dissociation from Rop more than a thousand-fold and forms a tight complex with nucleotide-free Rop. RopGEFs may represent the missing link in signal transduction from receptor kinases to Rops and their identification has implications for the evolution of the Rho molecular switch.

Rops appear to adopt the role of Ras and Rho proteins as pivotal regulators in plant signal transduction^{1,2}. As Rops are believed to relay the involved signals to downstream cellular targets through their GTP-bound form¹, Rop activation seems crucial in signal transduction. However, plants lack sequences homologous to classical RhoGEFs⁸, which contain a tandem arrangement of a catalytically active Dbl (diffuse B-cell lymphoma)-homology (DH) domain and a pleckstrin-homology (PH) domain⁷. Nucleotide dissociation from Rops is slow (see below), as in other small GTP-binding proteins (G proteins)⁵, so we reasoned that structurally different RhoGEFs should exist in plants to promote Rop activation. Precedents for RhoGEFs without the DH-PH motif exist, such as the highly active RhoGEF SopE^{9,10} from *Salmonella typhimurium* and the Dock180-related proteins^{11,12} with a conserved Dock-homology-region-2 (DHR2) that mediates interaction with the G protein and promotes guanine nucleotide exchange. ELMO proteins are critical regulators of Dock180, and a Dock180–ELMO complex appears to function as a two-part GEF for Rac¹¹. A single homologue of Dock180, named SPIKE1 (ref. 13) was identified in *Arabidopsis*. However, its role as GEF has not been confirmed.

To identify plant RopGEF candidate proteins, we applied the yeast two-hybrid system using the mutant Rop4(D121N) as bait. Previously, we have shown that the homologous mutation (D119N) in the nucleotide binding motif of Ras leads to a strongly decreased nucleotide affinity and increased GEF affinity while Ras(D119N) can still bind to downstream effectors^{14,15}, making small G proteins with the homologous mutation an ideal bait for both GEF and effector preys.

In our two-hybrid assay, the C-terminal part of SPIKE1 including

the DHR2 (SpkDHR2) failed to interact with Rop4(D121N), although the yeast cells grew well without reporter gene selection (Fig. 1a) and expressed both the bait and prey proteins (Supplementary Fig. 1). Using an *Arabidopsis* inflorescence library¹⁶ instead, several clones were identified that supported growth on selective media via the activation of both the *HIS3* and the *ADE2* reporter genes. Two of these, designated HM2.4 and HM19, displayed strong interaction with Rop4(D121N) (Fig. 1a) and were homologous to each other. Sequence analysis of the rescued clones revealed that the encoded proteins were amino-terminally deleted but contained the Pfam database¹⁷ profile PF03759 for a domain of unknown function (DUF315) present in a family of plant hypothetical proteins. Because we would expect RopGEFs to constitute a protein family with several members, in analogy to the mammalian Rho system⁷, we considered DUF315-containing proteins rather than the single-copy gene product SPIKE1 to be good RopGEF candidates. Nevertheless, the

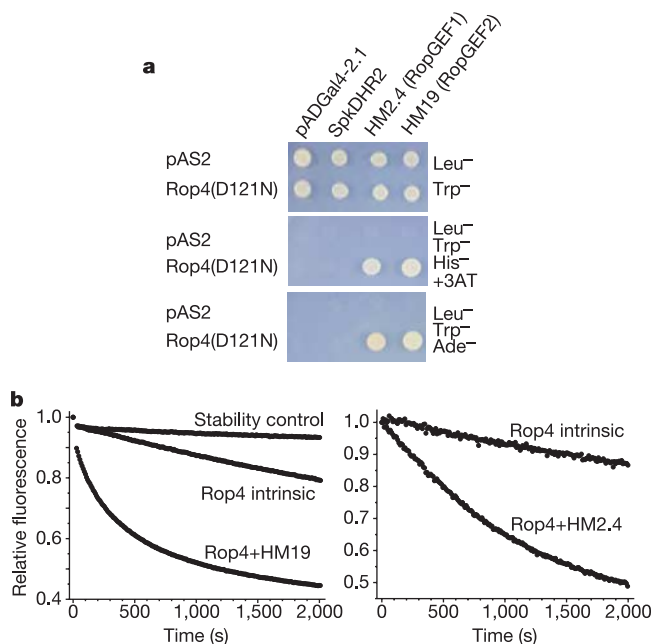


Figure 1 | Identification of RopGEFs from *Arabidopsis thaliana*. **a**, Yeast two-hybrid interactions of Rop4(D121N) with the RopGEFs HM2.4 and HM19 or SpkDHR2 (pAS2: BD-fusion vector; pADGal4-2.1: AD-fusion vector). Cells were selected for plasmids on Leu⁻ Trp⁻ medium. High-stringency selective plates (Leu⁻ Trp⁻ His⁻ with 10 mM 3AT or Leu⁻ Trp⁻ Ade⁻) were used to grow interacting clones. **b**, GEF activity of HM19 (2 μ M) and HM2.4 (2.9 μ M) in guanine nucleotide exchange assays with Rop4-mGDP (500 nM and 400 nM, respectively) in the presence of 5 mM MgCl₂ and 460-fold molar excess of unlabelled GDP. Stability control: without unlabelled GDP.

¹Max Planck Institute of Molecular Physiology, Otto-Hahn-Strasse 11, 44227 Dortmund, Germany.

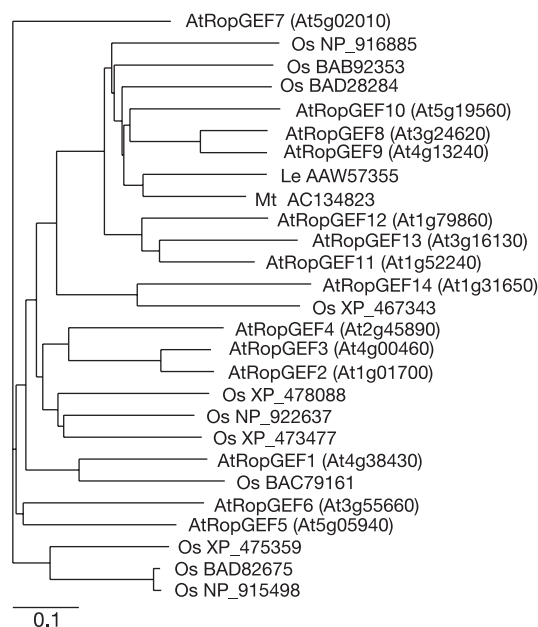


Figure 2 | The plant RopGEF family. Phylogenetic tree of sequences obtained from GenBank using BLAST (<http://www.ncbi.nlm.nih.gov/BLAST>). Phylogram was constructed with TreeView (<http://taxonomy.zoology.gla.ac.uk/rod/treeview.html>) after alignment with ClustalW (<http://www.ebi.ac.uk/clustalw>). *Arabidopsis thaliana* (At) RopGEFs are given with their corresponding loci. The accession numbers for rice (Os: *Oryza sativa*), the model legume 'barrel medic' (Mt: *Medicago truncatula*) and tomato (Le: *Lycopersicon esculentum*) are as indicated.

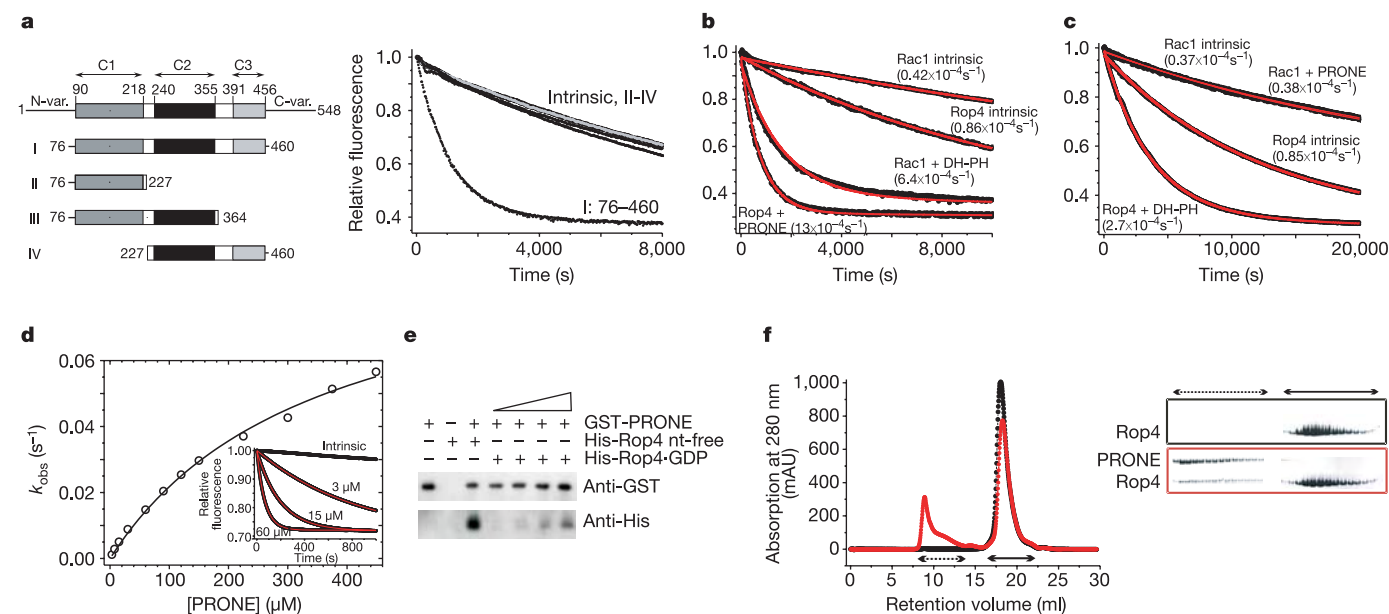


Figure 3 | Functional characterization of PRONE as genuine GEF domain. **a**, GEF activity of fragments I–IV, containing different combinations of conserved regions (C1–C3) from RopGEF1 in assays with Rop4. **b**, Comparison of GEF activity between PRONE and DH-PH. Rate constants (k_{obs}) are shown in parentheses, and fits over 40,000 s as red lines. **c**, Cross-reactive GEF activity of PRONE towards Rac1 and DH-PH towards Rop4 (fits over 60,000 s). **d**, Stopped-flow kinetic analysis. k_{obs} values were obtained with 150 nM Rop4 by single-exponential fits, shown in the inset for three different PRONE concentrations. Data points as means of three

existence of several ELMO-related homologues in the *Arabidopsis* genome may still argue for SPIKE1–ELMO complexes in plants, and their involvement in Rop activation cannot yet be excluded completely.

To determine whether the identified proteins HM19 and HM2.4 could be RopGEFs, we purified the proteins as recombinant GST fusion proteins and determined their GEF activity by measuring the release of fluorescently labelled mGDP (see Methods) from Rop4 in real time (Fig. 1b). Both proteins displayed significant GEF activity on Rop4, and thus were renamed RopGEF1 (HM2.4) and RopGEF2 (HM19).

Further database searches revealed that the full-length proteins belong to a family of highly related proteins (Fig. 2) with at least fourteen members in *Arabidopsis*, eleven orthologues in rice (*Oryza sativa*), one in tomato (*Lycopersicon esculentum*), and a sequence in *Medicago truncatula*. The plant proteins share 25–97% overall sequence identity but no similarity to any known RhoGEF of the DH-PH-type, to SopE, or the Dock180-related proteins. As shown with representatives from each species (Supplementary Fig. 2), they display variable amino and carboxy termini and a highly conserved central part composed of three regions (C1, C2 and C3) with considerable sequence homology. Apart from the conserved central part, they contain no other discernible motifs, and no sequence homologues were recognized in non-plant species.

On the basis of the high sequence conservation in the central part, we asked whether this portion in RopGEFs is sufficient for catalysis. RopGEF1 proved to be more stable than RopGEF2, so we used it for further investigations. To define the catalytic core element in RopGEF1, we expressed the central part (fragment I: amino acids 76–460) which showed high RopGEF activity (Fig. 3a). By comparison, Ras superfamily GEF domains⁶ such as Sec7, Cdc25 and DH are much shorter and we wondered whether the active GEF domain in RopGEF1 is shorter than the examined 385-residue fragment.

measurements were fitted to a hyperbolic function. **e**, Preferential binding of PRONE to nucleotide (nt)-free Rop4. Western blots of pull-downs with GST-tagged PRONE and His-tagged Rop4. The triangle indicates increasing sample volumes with Rop4-GDP, loaded to detect binding to PRONE. **f**, Complex formation in solution between PRONE and Rop4 as detected by gel filtration of Rop4 alone (black) or in the presence of PRONE (red). Peak fractions were analysed by SDS-PAGE and Coomassie-Blue staining. mAU, milli-absorbance units.

However, further deletion of the construct from the N or C terminus (fragments II to IV) completely inactivated the protein, arguing that the three regions (C1–C3) with high amino-acid conservation are necessary and sufficient for GEF activity. Thus the active, conserved domain was named PRONE, for plant-specific Rop nucleotide exchanger.

To evaluate the effectiveness of the plant RopGEFs, the PRONE domain of RopGEF1 was compared to a DH-PH construct of the mammalian Rac-specific GEF Tiam1 (T-lymphoma invasion and metastasis factor 1)¹⁸. The intrinsic dissociation of Rop4•mGDP was somewhat faster than that of human Rac1 (Fig. 3b). However, it was stimulated 15-fold by PRONE and thus is comparable to the 15-fold accelerated nucleotide release induced by DH-PH under the same conditions. Considering the high homology between plant Rops and mammalian Rho, especially Rac^{1,2,3,8}, we wondered about the cross-reactivity between the plant and mammalian components. While DH-PH was somewhat active on Rop4 with a threefold stimulation of the intrinsic nucleotide dissociation, PRONE was inactive on Rac1 under identical conditions (Fig. 3c). Likewise, no activity of PRONE

was observed towards mammalian Cdc42, RhoA or Ras (not shown). The cross-reactivity of DH-PH towards Rop4 may be explained by the identity of residues between Rac and Rop that were shown in structural¹⁹ and mutational analyses²⁰ to determine specific recognition of DH-PH-type GEFs. Because PRONE does not act on Rac1 despite a strong similarity to Rop, it apparently uses, apart from a different fold, a different set of interactions and possibly a different mode of action.

To determine the maximum GEF-induced nucleotide dissociation rate from Rop4, dissociation rate constants were measured with increasing concentrations of PRONE in a stopped-flow experiment (Fig. 3d). Although we did not reach complete saturation at 450 μ M PRONE, the data could be fitted to obtain the rate constant $k_{\max}$ of $9.8 \times 10^{-2} \text{ s}^{-1}$ corresponding to a more than thousand-fold stimulation of nucleotide dissociation. Comparable activations have been reported for the Tiam1-induced catalysis with the Rac isoforms, Rac1 (200-fold) and Rac2 (6,000-fold)¹⁸.

GEFs can be distinguished from other interaction partners by their ability to bind to the nucleotide-free state of the G protein, because catalysis of nucleotide exchange proceeds from a ternary G protein•nucleotide•GEF to a binary G protein•GEF complex and reverse⁶. PRONE did indeed preferentially interact with nucleotide-free Rop4 in a pull-down experiment (Fig. 3e). A nucleotide-free Rop4•PRONE complex could also be isolated by gel filtration chromatography (Fig. 3f). Taking these results together, and considering the data from the two-hybrid analysis, we regard the PRONE domain as a genuine GEF domain.

In view of the presence of multiple Rops and RopGEFs in *Arabidopsis*, it is unclear whether there is any specificity in the activation pattern and how it would be achieved. *Arabidopsis* contains eleven highly similar Rops that were classified into four subgroups (I–IV) with supposedly distinct functions³. We compared the GEF activity of PRONE towards Rop proteins of group II (Rop10), group III (Rop7) and group IV (Rop3 and Rop4) (Rop8 of group I was not investigated owing to its insolubility) and found that it was active on every Rop tested (Fig. 4a). However, the intrinsic nucleotide dissociation from Rop7 was comparatively fast, so the PRONE-stimulated acceleration was only twofold, while six- to 13-fold stimulations were obtained with Rop3, Rop4 and Rop10.

The spatial expression pattern of RopGEFs may influence the activation of certain Rop-dependent signal transduction pathways, at least for RopGEF1, because the expression level varies in different *Arabidopsis* tissues, with high levels in flowers and low expression in leaves, roots and siliques (Fig. 4b). Specific Rop activation may also be controlled by a different temporal expression pattern or by the participation of a particular membrane environment. With Rac and Tiam1 it has been shown that their functional interaction is favoured in the environment of the membrane²¹ where nucleotide exchange occurs *in vivo*. Furthermore, while we were preparing this manuscript, the RopGEF homologue from tomato was described as kinase partner protein (KPP), which is a peripheral membrane protein interacting with the receptor kinases LePRK1 and LePRK2, and is phosphorylated in pollen²². While the function of KPP remains unclear in this study, the data add significance to the notion that RopGEF activity could be controlled at the membrane by interaction with or phosphorylation by receptor kinases. Moreover, comparable phenotypes of Rop and KPP over-expressing pollen tubes²² suggest an *in vivo* role of RopGEFs as a bridge between receptor kinases and Rops.

The identification of a plant-specific family of RhoGEFs raises some interesting questions concerning the evolution of G proteins. GEFs, GTPase activating proteins (GAPs) and effectors for Rho proteins are highly conserved from fungi to animals. Plants contain, in addition to Rops, RopGAPs homologous to RhoGAPs and effector molecules with a CRIB (Cdc42/Rac interactive binding) domain homologous to that of mammalian effectors^{1–3,8}. The fact that

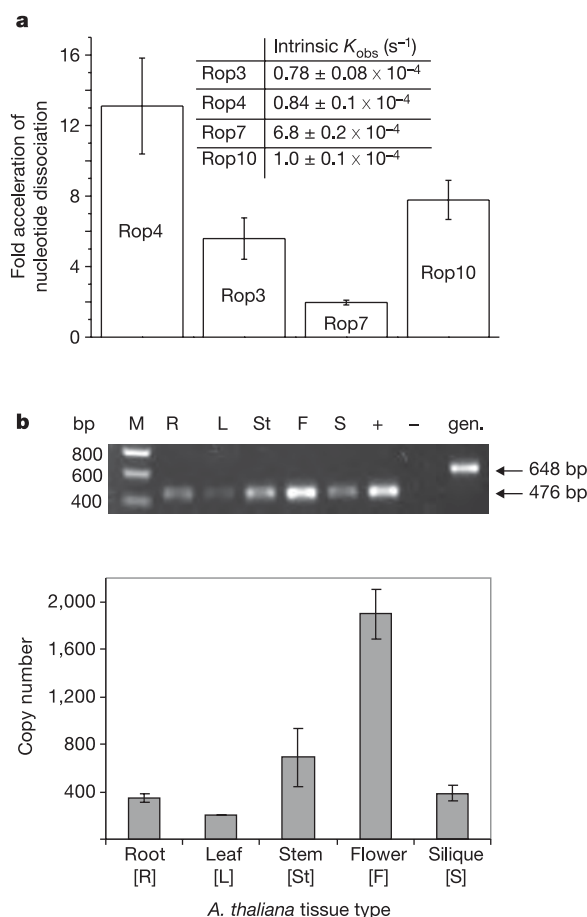


Figure 4 | RopGEF1 expression and specificity for Rops. **a**, GEF activity of PRONE of RopGEF1 towards different Rops. Magnitudes as means of three measurements with error bars representing sample standard deviations. Intrinsic k_{obs} values are given with standard deviation. **b**, Tissue distribution of RopGEF1 mRNA as analysed by copy numbers determined by quantitative real-time RT-PCR. Data represent means and standard deviations of three amplifications. Specific amplification of a 476-base-pair (bp) fragment of RopGEF1 cDNA was monitored by electrophoresis and ethidium bromide staining. Controls included pADGal4-2.1-HM2.4 (+) and pADGal4-2.1 (-). The purity of RNA preparations was established by the absence of a 648-bp intron-spanning fragment resulting from genomic DNA (gen.). M is the DNA molecular weight standard.

PRONE- and DH-PH-GEFs are not conserved between plants and opisthokonts²³ (fungi and animals) argues for a late addition of the GEFs to the repertoire of Rho regulation. This seems only to be true for the Rho branch of GTP-binding proteins because a genome-wide analysis of the small GTPase gene superfamily in *Arabidopsis* identified candidate GEFs for Rab, Arf and Ran homologous to Vps9/Rabex5, Sec7 and RCC1 (ref. 8). Nevertheless, a gene loss for DH-PH-GEFs in plants cannot be finally dismissed.

METHODS

Complementary DNA clones. cDNAs for Rop3, 8 and 10 were isolated by polymerase chain reaction (PCR) from a cDNA library¹⁶ or from reverse-transcribed (Superscript II, Invitrogen) total RNA (RNeasy Kit, Qiagen) from *Arabidopsis* inflorescences and inserted into pCR2.1 by TA-cloning (Invitrogen). Rop7 cDNA and pACT2-996-DHR2 encoding residues 996–1830 of SPIKE1 (SpkDHR2) were provided by D. Szymanski, pET28a(+)-Rop4 was a gift from K. Palme, and pGEX4T1-Tiam1DH-PH¹⁸ was a gift from R. Ahmadian. DNA inserts for additional plasmid constructs were amplified by PCR with primers linked to restriction enzyme sites suitable for cloning.

Protein expression and purification. RopGEF constructs and Rops lacking the C-terminal prenylation motif (Rop3: residues 1–180; Rop4: 1–180; Rop7: 1–180; Rop8: 1–192; Rop10: 1–182) were cloned as N-terminal glutathione-S-transferase (GST) fusion proteins in pGEX6P1 or pGEX4T1 (Amersham Biosciences). GST-tagged proteins were expressed in *Escherichia coli* BL21-Codon-Plus(DE3)-RIL (Stratagene) at 30 °C for 4 h after induction with 0.3 mM isopropyl-β-D-thiogalactopyranoside (IPTG). His-tagged full-length Rop4 was expressed from pET28a(+)-Rop4 overnight at 20 °C. Fusion proteins were purified from soluble fractions of cell lysates by glutathione (GSH-Sepharose 4 Fast Flow, Amersham Biosciences) or nickel (Ni-NTA Superflow, Qiagen) affinity chromatography. GST tags were removed, where applicable, by on-column digestion with PreScission protease (Amersham Biosciences) or thrombin (Serva). Eluted proteins were subjected to gel filtration (Superdex 75 or 200, Amersham Biosciences) and concentrated by ultrafiltration (Amicon, Millipore).

Yeast techniques. Yeast techniques and two-hybrid methods were performed according to the yeast protocols handbook and the Matchmaker GAL4 Two-Hybrid System 3 manual (Clontech) with *Saccharomyces cerevisiae* AH109. Rop4 (residues 1–180) with the D121N mutation generated by overlap-extension PCR²⁴ was cloned in pAS2²⁵ generating a fusion with the Gal4 DNA-binding domain (BD). Using pAS2-Rop4(D121N) as bait, we screened 1.23×10^5 prey clones of a two-hybrid library from *Arabidopsis* inflorescences in pADGal4-2.1¹⁶ as fusions with the Gal4 activation domain (AD). Cells were selected on medium lacking leucine (Leu[−]), tryptophan (Trp[−]) and histidine (His[−]) supplemented with 2.5 mM 3-amino-triazole (3AT). Interacting RopGEF clones comprised residues 36–548 of RopGEF1 (HM2.4) and 39–485 of RopGEF2 (HM19). Interactions were further tested by spotting 10^5 cells co-transformed with pAS2-Rop4(D121N) and pADGAL4-2.1-HM19, pADGAL4-2.1-HM2.4 and pACT2-996-DHR, respectively, on Leu[−] Trp[−] His[−] plates with 10 mM 3AT and Leu[−] Trp[−] plates without adenine (Ade[−]). Identical results were obtained for several co-transformants. Expression of GAL4 BD- or AD-fusions in AH109 was verified by immunoblotting of protein extracts with GAL4 BD (Santa Cruz) or GAL4 AD (Clontech) antibodies and anti-mouse horseradish peroxidase conjugate (Sigma).

Guanine nucleotide exchange assay (GEF activity assay). Intrinsic and GEF-stimulated dissociation of N-methylanthraniloyl [mant]-GDP (mGDP) from purified G proteins was monitored on a Spex FluoroMax fluorometer (Jobin Yvon) and by stopped-flow analysis (SX.18MV, Applied Photophysics) as described before²⁶ with excitation at 360 nm and emission at 440 nm. Reactions were carried out at 20 °C in 50 mM Tris/HCl pH 7.3, 100 mM NaCl, 3 mM β-mercaptoethanol, 10 mM MgCl₂ with 200 nM G protein, 2 μM GEF and 460-fold excess of unlabelled GDP unless otherwise mentioned. RopGEF constructs were used as GST-fusions as they were more stable and tag cleavage did not alter activity. Stopped-flow data were fitted to a single-exponential decay function with the program Origin7 (OriginLab Corporation). The function for fitting data of all other exchange assays contained an additional linear portion.

Pull-down assays and complex formation. Assays contained 100 μg of GST-fused PRONE immobilized on GSH-Sepharose beads and 25 μg His-tagged Rop4-GDP or nucleotide-free²⁷ Rop4 in 30 mM Tris/HCl pH 8.0, 100 mM NaCl, 3 mM β-mercaptoethanol, 5% (v/v) glycerol with 100 μg bovine serum albumin as a non-specific competitor. After 1 h incubation (4 °C), samples were washed five times in the presence of 300 mM NaCl and eluted by boiling in

Laemmli-buffer²⁸. After electrophoresis, samples were blotted and probed with anti-GST or anti-His (Qiagen) antibodies and anti-mouse horseradish peroxidase conjugate. For complex formation, PRONE was incubated (20 min, room temperature) with fivefold molar excess of Rop4-GDP in the presence of 20 mM EDTA in 50 mM Tris/HCl pH 7.4, 300 mM NaCl and 3 mM β-mercaptoethanol. The sample was subsequently subjected to gel filtration (S200, 10/30, Amersham Biosciences) with 50 mM Tris/HCl pH 7.3, 300 mM NaCl, 3 mM β-mercaptoethanol, 10% (v/v) glycerol, and 5 mM EDTA. Fractions were analysed by SDS-PAGE and Coomassie-Blue staining. The absence of nucleotide in the complex was checked by HPLC²⁹.

Quantitative real-time reverse-transcription PCR. 240 ng total RNA (RNeasy kit, Qiagen) from tissues of *A. thaliana* (ecotype Columbia) was reverse-transcribed (Superscript II, Invitrogen) into cDNA by Oligo dT priming. RopGEF1 cDNA was amplified from 1/50 of the reverse-transcription assay or, as control from genomic DNA (DNeasy kit, Qiagen), with forward (5′-CGCAA TTCGATGTGGTCTGGGATGACAAACAAC-3′) and reverse (5′-CGCCTCGA GTTAATCACCAGAAGACTATGCC-3′) primers on a LightCycler System (Roche) using the QuantiTect SYBR green PCR kit (Qiagen). Samples were normalized to total RNA as the preferred method³⁰. Data were analysed using a standard curve with the Second Derivative Maximum method (LightCycler Operator's Manual, Roche).

Received 28 March; accepted 3 June 2005.

Published online 26 June 2005.

- Zheng, Z. & Yang, Z. The Rop GTPase: an emerging signalling switch in plants. *Plant Mol. Biol.* **44**, 1–9 (2000).
- Valster, A. H., Hepler, P. K. & Chernoff, J. Plant GTPases: the Rhos in bloom. *Trends Cell Biol.* **10**, 141–146 (2000).
- Yang, Z. Small GTPases: versatile signalling switches in plants. *Plant Cell*, S375–S388 (2002).
- Bourne, H. R., Sanders, D. A. & McCormick, F. The GTPase superfamily: a conserved switch for diverse cell functions. *Nature* **348**, 125–132 (1990).
- Bourne, H. R., Sanders, D. A. & McCormick, F. The GTPase superfamily: conserved structure and molecular mechanism. *Nature* **349**, 117–126 (1991).
- Cherfils, J. & Chardin, P. GEFs: structural basis for their activation of small GTP-binding proteins. *Trends Biochem. Sci.* **24**, 306–311 (1999).
- Rossman, K. L., Der, C. J. & Sondek, J. GEF means go: turning on the Rho GTPases with guanine nucleotide-exchange factors. *Nature Rev. Mol. Cell Biol.* **6**, 167–180 (2005).
- Vernoud, V., Horton, A. C., Yang, Z. & Nielsen, E. Analysis of the small GTPase gene superfamily of *Arabidopsis*. *Plant Physiol.* **131**, 1191–1208 (2003).
- Hardt, W. D., Chen, L. M., Schuebel, K. E., Bustelo, X. R. & Galán, J. E. S. *typhimurium* encodes an activator of Rho GTPases that induces membrane ruffling and nuclear responses in the host cell. *Cell* **93**, 815–826 (1998).
- Rudolph, M. G. *et al.* Biochemical analysis of SopE from *Salmonella typhimurium*, a highly efficient guanosine nucleotide exchange factor for RhoGTPases. *J. Biol. Chem.* **274**, 30501–30509 (1999).
- Brugnera, E. *et al.* Unconventional Rac-GEF activity is mediated through the Dock180-ELMO complex. *Nature Cell Biol.* **4**, 574–582 (2002).
- Meller, N., Irani-Tehrani, M., Kiosses, W. B., Del Pozo, M. A. & Schwartz, M. A. Zizimin1, a novel Cdc42 activator, reveals a new GEF domain for Rho proteins. *Nature Cell Biol.* **4**, 639–647 (2002).
- Qiu, J. L., Jilk, R., Marks, M. D. & Szymanski, D. B. The *Arabidopsis* SPIKE1 gene is required for normal cell shape and tissue development. *Plant Cell* **14**, 101–118 (2002).
- Schmidt, G. *et al.* Biochemical and biological consequences of changing the specificity of p21(ras) from guanosine to xanthosine nucleotides. *Oncogene* **12**, 87–96 (1996).
- Cool, R. H. *et al.* The Ras mutant D119N is both dominant negative and active. *Mol. Cell Biol.* **19**, 6297–6305 (1999).
- Fan, H. Y., Hu, Y., Tudor, M. & Ma, H. Specific interactions between the K domains of AG and AGLs, members of the MADS domain family of DNA binding proteins. *Plant J.* **12**, 999–1010 (1997).
- Bateman, A. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **32**, D138–D141 (2004).
- Haeusler, L. C., Blumenstein, L., Stege, P., Dvorsky, R. & Ahmadian, M. R. Comparative functional analysis of the Rac GTPases. *FEBS Lett.* **55**, 556–560 (2003).
- Worthylake, D. K., Rossman, K. L. & Sondek, J. Crystal structure of Rac1 in complex with the guanine nucleotide exchange region on Tiam1. *Nature* **408**, 682–688 (2000).
- Karnoub, A. E. *et al.* Molecular basis for Rac1 recognition by guanine nucleotide exchange factors. *Nature Struct. Biol.* **8**, 1037–1041 (2001).
- Robbe, K., Otto-Bruc, A., Chardin, P. & Antonny, B. Dissociation of GDP dissociation inhibitor and membrane translocation are required for efficient activation of Rac by the Dbl homology-pleckstrin homology region of Tiam. *J. Biol. Chem.* **278**, 4756–4762 (2003).
- Kaethien, P. *et al.* Kinase partner protein interacts with the LePRK1 and LePRK2

- receptor kinases and plays a role in polarized pollen tube growth. *Plant J.* **42**, 492–503 (2005).
23. Baldauf, S. L. The deep roots of eukaryotes. *Science* **300**, 1703–1706 (2003).
24. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K. & Pease, L. R. Site-directed mutagenesis by overlap-extension using the polymerase chain reaction. *Gene* **77**, 51–59 (1989).
25. Harper, J. W., Adami, G. R., Wei, N., Keyomarsi, K. & Elledge, S. J. The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1 cyclin-dependent kinases. *Cell* **75**, 805–816 (1993).
26. Lenzen, C., Cool, R. H., Prinz, H., Kuhlmann, J. & Wittinghofer, A. Kinetic analysis by fluorescence of the interaction between Ras and the catalytic domain of the guanine nucleotide exchange factor Cdc25Mm. *Biochemistry* **37**, 7420–7430 (1998).
27. John, J. *et al.* Kinetics of interaction of nucleotides with nucleotide-free H-ras p21. *Biochemistry* **29**, 6058–6065 (1990).
28. Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685 (1970).
29. Lenzen, C., Cool, R. H. & Wittinghofer, A. Analysis of intrinsic and Cdc25-stimulated guanine nucleotide exchange of p21ras-nucleotide complexes by fluorescence measurements. *Methods Enzymol.* **255**, 95–109 (1995).
30. Bustin, S. A. Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *J. Mol. Endocrinol.* **29**, 23–39 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Koerner, T. Mitroshina and D. Kuehlmann for technical assistance; R. Ahmadian for purified Rac1, Cdc42, RhoA, Anti-GST antibodies and for helpful discussions; The *Arabidopsis* Biological Resource Center for the two-hybrid cDNA library (CD4-30) and pAS2; E. W. Weiler for plant material; K. Unfried and J. Abel for providing the LightCycler, and K. Bierhals for critical reading of the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft. C.T. thanks the Boehringer Ingelheim Fonds for support.

Author Information Accession numbers for the plant RopGEF proteins in *Arabidopsis thaliana*, *Oryza sativa*, *Medicago truncatula* and *Lycopersicon esculentum* are listed in Fig. 2. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.B. (antje.berken@mpi-dortmund.mpg.de).

Hassall's corpuscles instruct dendritic cells to induce CD4⁺CD25⁺ regulatory T cells in human thymus

Norihiko Watanabe¹, Yi-Hong Wang¹, Heung Kyu Lee¹, Tomoki Ito¹, Yui-Hsi Wang¹, Wei Cao¹ & Yong-Jun Liu¹

Hassall's corpuscles—first described in the human thymus over 150 years ago¹—are groups of epithelial cells within the thymic medulla. The physical nature of these structures differs between mammalian species². Although Hassall's corpuscles have been proposed to act in both the removal of apoptotic thymocytes^{3,4} and the maturation of developing thymocytes⁵ within the thymus, the function of Hassall's corpuscles has remained an enigma. Here we report that human Hassall's corpuscles express thymic stromal lymphopoietin (TSLP). Human TSLP activates thymic CD11c-positive dendritic cells to express high levels of CD80 and CD86. These TSLP-conditioned dendritic cells are then able to induce the proliferation and differentiation of CD4⁺CD8[−]CD25[−] thymic T cells into CD4⁺CD25⁺FOXP3⁺ (forkhead box P3) regulatory T cells. This induction depends on peptide-major histocompatibility complex class II interactions, and the presence of CD80 and CD86, as well as interleukin 2. Immunohistochemistry studies reveal that CD25⁺CTLA4⁺ (cytotoxic T-lymphocyte-associated protein 4) regulatory T cells associate in the thymic medulla with activated or mature dendritic cells and TSLP-expressing Hassall's corpuscles. These findings suggest that Hassall's corpuscles have a critical role in dendritic-cell-mediated secondary positive selection of medium-to-high affinity self-reactive T cells, leading to the generation of CD4⁺CD25⁺ regulatory T cells within the thymus.

Hassall's corpuscles, the corpuscular bodies of epithelial cells located within the thymic medulla, were first described by Arthur Hill Hassall in 1849 (ref. 1). Hassall's corpuscles are well developed in the thymus of humans and guinea-pigs, but poorly developed in the thymus of mice and rats². Hassall's corpuscles were suggested to represent the 'graveyard' for dead thymocytes^{3,4}, and also the 'privileged' area for the maturation of medullary thymocytes⁵. Other studies have provided evidence suggesting that Hassall's corpuscles are active in cytokine or growth factor receptor-mediated cell signalling, transcription, and metabolism⁶. Hassall's corpuscles were also found to express cytokines such as interleukin (IL)-7, transforming growth factor (TGF)- α , CD30 ligand, stromal-cell-derived factor 1 (SDF-1) and macrophage-derived chemokine (MDC)^{7–11}. These data suggest that Hassall's corpuscles may actively communicate with developing T cells and antigen-presenting cells within the thymus.

Our interest in human Hassall's corpuscles stemmed from the observation that epithelial cells within these structures express TSLP, an IL-7-like cytokine (ref. 12 and Fig. 1a). Both IL-7 and TSLP have major species differences between humans and mice. Unlike murine IL-7 and TSLP, which are critical for murine B- and T-cell development, human IL-7 is critical for T-cell but not B-cell development, and human TSLP activates only dendritic cells^{12–14}. Human TSLP

strongly activates CD11c-positive immature myeloid dendritic cells to upregulate major histocompatibility complex (MHC) class II molecules, dendritic cell lysosomal-associated membrane protein (DC-LAMP, which is only expressed by activated dendritic cells) and the co-stimulatory molecules CD80 and CD86 (refs 12, 14). In the human thymus, TSLP was expressed by the epithelial cells of Hassall's corpuscles in the medulla of the thymus (Fig. 1a). The staining of this antibody was TSLP-specific as recombinant TSLP, but not recombinant IL-7, completely blocked the staining, and the isotype control did not give any positive staining (Supplementary Fig. S1). The expression of TSLP by the epithelial cells of Hassall's corpuscles in human thymus is supported by the additional observations that TSLP was originally cloned from thymic epithelial cells¹⁵, and TSLP messenger RNA was found to be expressed by epithelial cells but not by haematopoietic cells¹⁴. The thymic medulla is localized by the presence of CD11c-positive medullary dendritic cells and Hassall's corpuscles (Supplementary Fig. S2). We found that TSLP-expressing Hassall's corpuscles co-localized with CD11c⁺DC-LAMP⁺ activated dendritic cells within the central part of thymic medulla, where DC-LAMP marks mature dendritic cells (Fig. 1a and Supplementary Fig. S2). In contrast, CD11c⁺DC-LAMP[−] immature dendritic cells were mainly scattered in the cortex and cortico-medullary junction of the thymus (Supplementary Fig. S2). Because immature dendritic cells within fetal and newborn thymuses are less likely to be exposed to Toll-like receptor-mediated activation by microbes, we proposed that Hassall's corpuscles are programmed to express TSLP in order to activate immature dendritic cells to become CD11c⁺DC-LAMP⁺ activated dendritic cells in the thymic medulla. Indeed, we found that TSLP induced both peripheral blood and thymic CD11c-positive immature dendritic cells to upregulate DC-LAMP, human leukocyte antigen (HLA)-DR, CD80 and CD86 (Fig. 1b and data not shown). TSLP also stimulated both peripheral blood and thymic CD11c-positive immature dendritic cells to produce high levels of the chemokines thymus- and activation-regulated chemokine (TARC) and MDC, but not the pro-inflammatory cytokines IL-12 and tumour-necrosis factor (TNF)- α (Fig. 1c)—a unique signature of TSLP-mediated activation of dendritic cells¹⁴.

We proposed that TSLP-activated dendritic cells (TSLP-DCs) within the thymic medulla might be critical for stimulating high-affinity autoreactive developing T cells to differentiate into CD4⁺CD25⁺ regulatory T (T_R) cells, because: (1) TSLP-DCs were shown to induce a strong homeostatic proliferation of peripheral naive CD4⁺ T cells¹², and homeostatic proliferation of CD4⁺ T cells in the periphery and positive selection of the developing CD4⁺ T cells in the thymus appear to be regulated by similar cellular and molecular mechanisms^{16–18}; (2) the development of CD4⁺CD25⁺ T_R cells requires CD28 signalling^{19,20}, and TSLP-DCs within the

¹Department of Immunology and Center for Cancer Immunology Research, The University of Texas, M. D. Anderson Cancer Center, 7455 Fannin Street, Houston, Texas 77054, USA.

thymic medulla may represent the only dendritic cells expressing the CD28 ligands CD80 and CD86; and (3) it has recently become evident that mature dendritic cells are capable of expanding the $CD4^+CD25^+$ peripheral T-cell population²¹. Therefore, we evaluated the consequences of culturing different subpopulations of human thymocytes with dendritic cells that had been activated with TSLP.

T-cell lineage thymocytes were isolated from human thymus and cultured for 7 days with dendritic cells precultured for 24 h with medium only, or medium containing TSLP, IL-7, CD40 ligand or polyinosinic:polycytidylic acid (poly(I:C)). We found that TSLP-DCs induced a vigorous expansion in the total number of thymocytes, whereas unstimulated dendritic cells, or dendritic cells exposed to IL-7, CD40 ligand or poly(I:C), did not (Supplementary Fig. S3a). However, phenotypic analyses showed that the expanded thymocytes were mainly $CD4^+CD8^-$ cells, and most expressed CD25 (Supplementary Fig. S3b–d). We then used a cell sorting method (see Methods) to separate T-cell lineage thymocytes into $CD4^-CD8^-$, $CD4^+CD8^+$, $CD4^+CD8^-$ and $CD4^-CD8^+$ populations. Approximately 4% of the $CD4^+CD8^-$ population expressed CD25 as previously reported (ref. 22 and Fig. 2a). After 7 days of culture with TSLP-DCs, there was more than a tenfold expansion of the $CD4^+CD25^+$ T cells within the $CD4^+CD8^-$ population (Fig. 2a). The other three populations did not expand in this culturing system (Fig. 2a). We concluded that TSLP-DCs could induce an increase in the number of $CD4^+CD25^+$ T cells in cultured $CD4^+CD8^-$ thymocytes.

We then investigated whether TSLP-DCs could induce the expansion of $CD4^+CD25^+$ T_R cells, or the differentiation of $CD4^+CD25^-$ cells into $CD4^+CD25^+$ T_R cells. To determine this, $CD4^+CD8^-$ thymocytes were separated into $CD4^+CD25^-$ and $CD4^+CD25^+$ subpopulations, labelled with carboxy-fluorescein diacetate succinimidyl ester (CFSE), and cultured with medium only, or in the presence of TSLP-DCs, unstimulated dendritic cells or IL-2 plus anti-CD3 and anti-CD28 antibodies. Whereas TSLP-DCs induced $CD4^+CD25^-$ cells to proliferate and differentiate into $CD4^+CD25^+$ T cells in large numbers, TSLP-DCs induced only the limited

proliferation and expansion of $CD4^+CD25^+$ T cells (Fig. 2b and Supplementary Fig. S4). In contrast, IL-2 plus anti-CD3 and anti-CD28 induced the strong proliferation and expansion of both $CD4^+CD25^+$ and $CD4^+CD25^-$ populations (Fig. 2b). To exclude the possibility that removing the IL-2-producing $CD4^+CD25^-$ cells from the $CD4^+CD25^+$ T_R cell population affected the selective expansion of $CD4^+CD25^+$ T_R cells by TSLP-DCs, we performed cell-mixing experiments with $CD4^+CD25^+$ T_R cells or CFSE-labelled $CD4^+CD25^-$ cells. Thymic-derived $CD4^+CD25^-$ cells were labelled with CFSE and mixed with a non-labelled $CD4^+CD25^+$ T_R cell population at a 1:9 $CD25^+CD25^-$ cell ratio. This ratio was chosen because the original $CD4^+CD8^-$ thymocyte population is a mixture of CD25-positive and CD25-negative cells at a 1:9 ratio (Fig. 2c, left). The cell populations were then cultured together with TSLP-DCs and the proliferative response of the two populations compared. We found that the $CD4^+CD25^-$ cells dominantly proliferated and became the principal CD25-positive cells in this mixture, but the $CD4^+CD25^+$ regulatory T cells underwent limited expansion and represented a minor population of CD25-positive cells (Fig. 2c). The predominant differentiation of $CD4^+CD25^-$ cells into CD25-positive cells was confirmed by mixed cultures of CFSE-labelled $CD4^+CD25^-$ regulatory T cells and non-labelled $CD4^+CD25^-$ cells with TSLP-DCs (Supplementary Fig. S5). Therefore, these data indicate that TSLP-DCs preferentially induce $CD4^+CD25^-$ thymocytes to differentiate into $CD4^+CD25^+$ T cells.

We next examined whether $CD4^+CD25^+$ T cells generated from $CD4^+CD8^-CD25^-$ thymocytes, when cultured with TSLP-DCs, exhibit the molecular signature of $CD4^+CD25^+$ T_R cells—*Foxp3* expression and suppressive function^{23,24}. We found that $CD4^+CD25^+$ T cells generated from culturing with TSLP-DCs, but not $CD4^+CD25^+$ T cells generated from culturing with anti-CD3 and anti-CD28 (with or without IL-2), expressed *Foxp3* mRNA at a level similar to the *Foxp3* mRNA level in freshly isolated thymic $CD4^+CD25^+$ T_R cells (Fig. 3a, b). Both the *in vivo*-derived $CD4^+CD25^+$ T_R cells and the *in vitro*-generated $CD4^+CD25^+$ T cells showed little response to anti-CD3 and anti-CD28 (Fig. 3c).

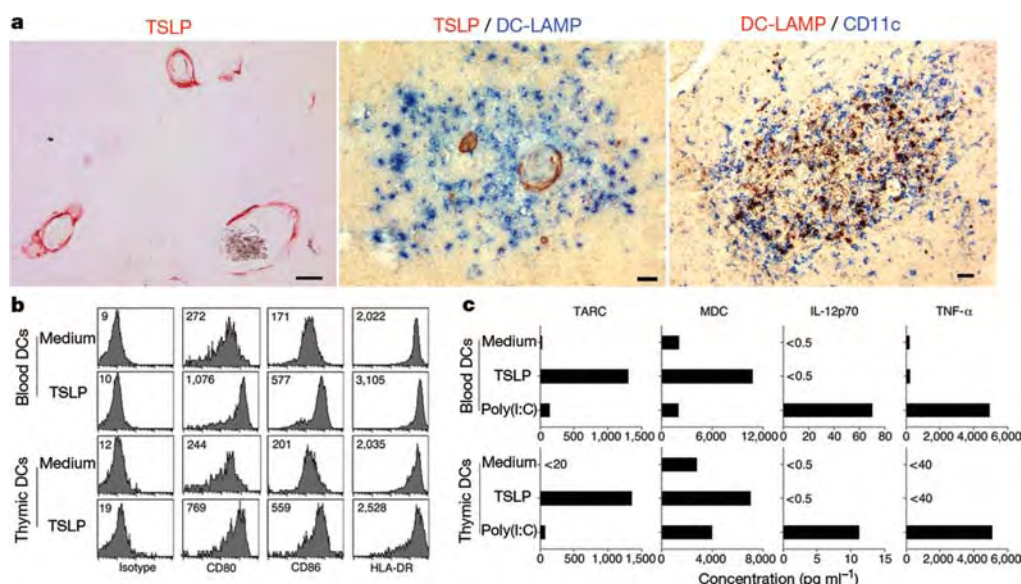


Figure 1 | Human TSLP activates thymic CD11c-positive dendritic cells. **a**, TSLP immunostaining showing that Hassall's corpuscles in the thymic medulla express TSLP (red) (left panel). Double staining of TSLP and DC-LAMP showing an association between TSLP-positive Hassall's corpuscles (red) and DC-LAMP-positive activated dendritic cells (blue) (middle panel). Double staining of DC-LAMP and CD11c showing two dendritic cell populations: CD11c⁺DC-LAMP⁺ activated dendritic cells (dark brown) and CD11c⁺DC-LAMP⁻ resting dendritic cells (blue) (right

panel). All scale bars, 50 μ m. **b**, **c**, Peripheral blood and thymic dendritic cells stimulated with TSLP or poly(I:C), or cultured in medium only. CD80, CD86 and HLA-DR expression as determined by flow cytometry (**b**). The numbers in the histograms indicate the mean fluorescence intensity. Production of TARC, MDC, IL-12p70 (the heterodimeric biologically active form of IL-12) and TNF- α by dendritic cells as determined by an enzyme-linked immunosorbent assay (**c**).

In contrast, $CD4^+CD25^-$ thymocytes vigorously proliferated in response to anti-CD3 and anti-CD28. This proliferation was inhibited by $CD4^+CD25^+$ T_R cells isolated from thymus, as well as $CD4^+CD25^+$ T cells generated from $CD4^+CD25^-$ thymocytes cultured with TSLP-DCs. These findings show that the newly generated $CD4^+CD25^+$ T cells possess some of the key features of suppressor T cells previously identified in peripheral T cells from mice and humans^{23,24}.

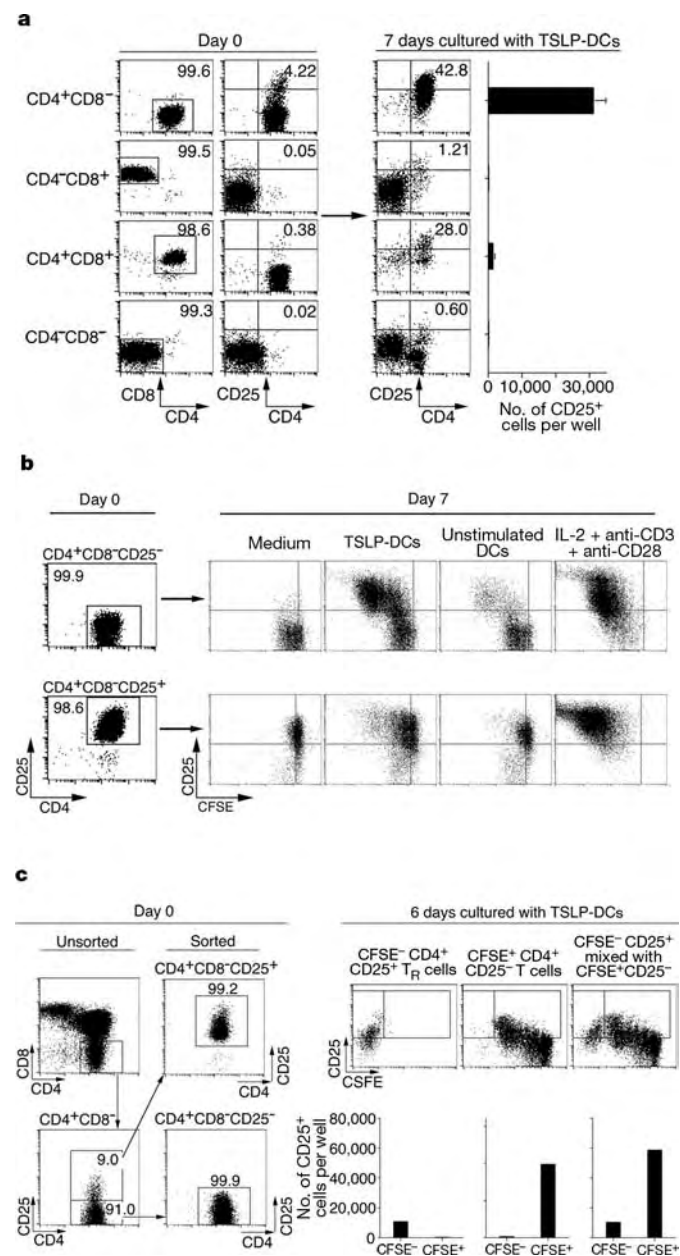


Figure 2 | TSLP-DCs induce the generation of $CD4^+CD8^-CD25^+$ T cells from $CD4^+CD8^-CD25^-$ thymocytes. **a**, Flow cytometric analysis of four thymocyte populations before and after culturing with TSLP-DCs. The numbers in the dot plots indicate the percentage of cells in that gate. The number of $CD25^+$ cells is shown in the bar graph as the mean and s.d. of triplicate wells. **b**, Flow cytometric analysis of CFSE-labelled $CD4^+CD8^-CD25^-$ and $CD4^+CD8^-CD25^+$ cells before and after culturing. **c**, Flow cytometric analysis of unsorted T -cell lineage thymocytes and two sorted $CD4^+CD8^-$ populations before culturing (left panel), and CFSE $^-CD4^+CD25^+$ cells, CFSE $^+CD4^+CD25^-$ cells and the mixture of two populations at an original 1:9 $CD25^+ : CD25^-$ cell ratio cultured with TSLP-DCs (right panel). The number of CFSE $^+/-CD25^+$ cells is shown in the bar graph.

As a first step towards identifying the molecules that enable TSLP-DCs to induce the generation of $CD4^+CD25^+$ T_R cells from $CD4^+CD8^-CD25^-$ thymocytes, blocking monoclonal antibodies to MHC class II (HLA-DR), CD80 plus CD86, and IL-2 all strongly inhibited the generation of $CD4^+CD25^+$ T_R cells (Fig. 3d and Supplementary Fig. S6). This is consistent with the previous findings regarding the critical role of self-peptide-MHC class II complexes, CD28 signalling through CD80 and CD86, and IL-2 in the generation of $CD4^+CD25^+$ T_R cells^{19,25-27}, and also the role of CD80 and CD86 in the expansion of peripheral $CD4^+CD25^+$ T_R cells by mature dendritic cells²⁰.

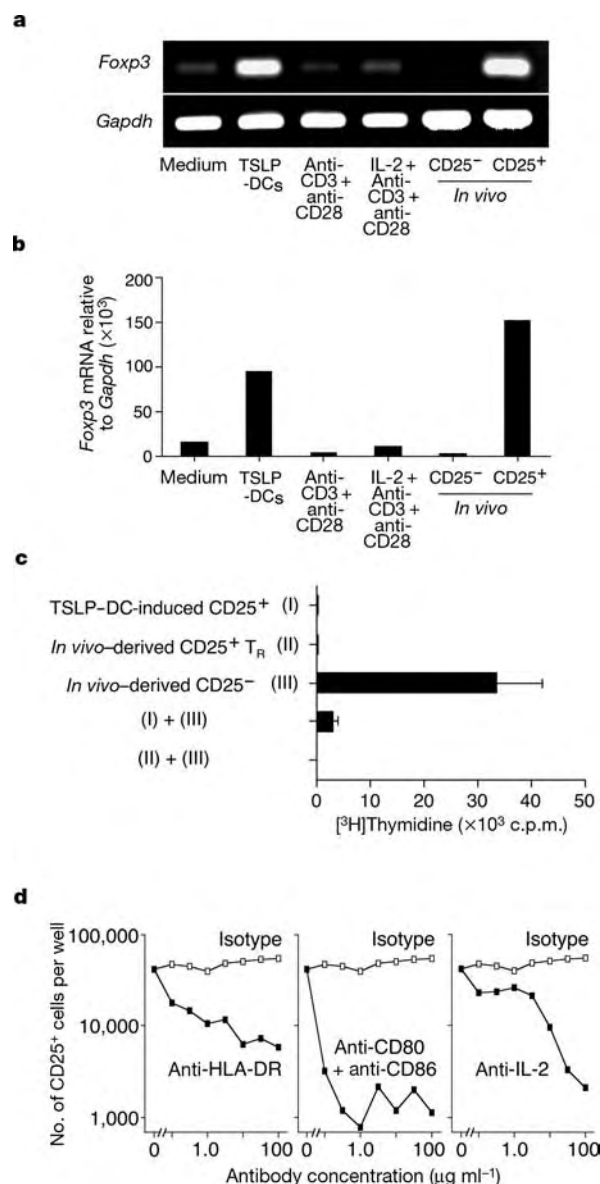


Figure 3 | TSLP-DC-induced $CD4^+CD25^+$ cells display features of $CD4^+CD25^+$ T_R cells. **a**, **b**, Conventional (**a**) and real-time quantitative (**b**) RT-PCR analysis for *Foxp3* mRNA expression by: $CD4^+CD25^-$ cells cultured with medium; $CD4^+CD25^+$ cells generated from $CD4^+CD25^-$ cells by the indicated conditions; and *in vivo*-derived thymic $CD4^+CD25^+$ T_R cells and $CD4^+CD25^-$ cells. **c**, Suppressive function assessed by [3H]thymidine incorporation. TSLP-DC-induced $CD4^+CD25^+$ T_R cells, *in vivo*-derived thymic $CD4^+CD25^+$ T_R cells and $CD4^+CD25^-$ cells, and the indicated mixtures of these populations, were re-stimulated by anti-CD3 and anti-CD28 antibodies. Error bars are s.d. of triplicate wells. **d**, Thymic $CD4^+CD25^-$ cells cultured with TSLP-DCs in the presence of neutralizing anti-HLA-DR, anti-CD80 plus anti-CD86, and anti-IL-2 (filled squares) antibodies, or isotype control (open squares) at various concentrations.

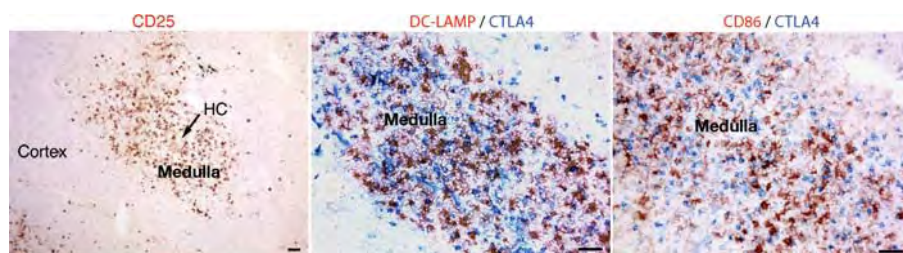


Figure 4 | $CD4^+CD25^+CTLA4^+$ T_R cells are associated with $DC-LAMP^+CD86^+$ activated dendritic cells within the thymic medulla. Immunostaining of CD25 (red) in the human thymus (left panel). Double

staining of DC-LAMP (red) and CTLA4 (blue) in the thymic medulla (middle panel). Double staining of CD86 (red) and CTLA4 (blue) in the thymic medulla (right panel). All scale bars, 50 μ m. HC, Hassall's corpuscle.

Because TSLP-DCs have the ability to induce peripheral allogeneic naive $CD4^+$ T cells to differentiate into T-helper (T_H) subtype 2 effector T cells¹⁴, the ability of TSLP-DCs to generate $CD4^+CD25^+$ T_R cells *de novo* from $CD4^+CD25^-$ precursors may be restricted to the $CD4^+CD8^-CD25^-$ single-positive thymocyte stage and not single-positive T cells from the periphery. To confirm this, naive $CD4^+CD25^-$ T cells from adult peripheral blood and $CD4^+CD8^-CD25^-$ thymocytes were each cultured with the same allogeneic TSLP-DCs for 7 days. The TSLP-DCs induced the peripheral blood naive $CD4^+$ T cells to undergo strong proliferation, which led to the generation of T_H2 cells that produced IL-4 and expressed CD25, but did not express *Foxp3* mRNA (Supplementary Fig. S7a–c). The TSLP-DCs also induced $CD4^+CD8^-CD25^-$ thymocytes to undergo strong proliferation; however, this led to the generation of T cells that did not produce T_H1 or T_H2 cytokines, but were CD25-positive and expressed *Foxp3* mRNA (Supplementary Fig. S7a–c). Furthermore, IL-2 plus anti-CD3 and anti-CD28 induced both peripheral blood naive $CD4^+$ T cells and $CD4^+CD8^-CD25^-$ thymocytes to undergo vigorous proliferation, which led to the generation of T cells that produced interferon- γ and expressed CD25, but did not express *Foxp3* mRNA (Supplementary Fig. S7a–c). These data suggest that TSLP-DCs only induce $CD4^+CD8^-CD25^-$ thymocytes, but not peripheral blood naive $CD4^+$ T cells, to differentiate into $CD4^+CD25^+Foxp3^+$ T_R cells.

We next examined the localization of $CD4^+CD25^+$ T_R cells within the thymus, and found that they were exclusively localized within the thymic medulla in close association with DC-LAMP⁺ or CD86⁺ activated dendritic cells, and Hassall's corpuscles (Fig. 4). These data provide anatomical proof that human $CD4^+CD25^+CTLA4^+$ T_R cells are generated in the thymic medulla, in close association with activated dendritic cells and Hassall's corpuscles.

Previous studies suggest that epithelial cells are critical for positive selection of the developing T cells in the thymic cortex, but that dendritic cells are critical for negative selection of high-affinity developing T cells in the thymic medulla^{28,29}. However, the finding that $CD4^+CD25^+$ T_R cells are positively selected according to their medium- to high-affinity binding to self-peptide–MHC complexes suggests that not all high-affinity T cells are deleted within the thymus^{25,26}. The observation that the development of $CD4^+CD25^+$ T_R cells depends on CD28 signalling¹⁹ suggests that dendritic cells expressing the ligands for CD28 (CD80 and CD86) may prevent the deletion of high-affinity thymocytes and induce their differentiation into $CD4^+CD25^+$ T_R cells. Our present study suggests that the epithelial cells within Hassall's corpuscles are developmentally programmed to express TSLP, which activates a subpopulation of dendritic cells in the thymic medulla to express CD80 and CD86. Although the immature dendritic cells within the cortico-medullary junction may be critical for negative selection, TSLP-DCs in the central part of the medulla may be critical for the positive selection of high-affinity autoreactive T cells to differentiate into $CD4^+CD25^+$ T_R cells. Therefore, we conclude that Hassall's corpuscles in the thymic medulla may have a critical role in the coordination of dendritic-cell-mediated central tolerance. This tolerance is not

achieved through the clonal deletion of self-reactive thymocytes, but by means of converting self-reactive thymocytes into anergic and suppressive regulatory T cells.

METHODS

Dendritic cell purification and culture. This study was approved by the institutional review board for human research at the M. D. Anderson Cancer Center. Human thymuses from fetuses (19–23 weeks of gestation), newborns and children (2 days to 2 years old) were obtained from Advanced Bioscience Resources and the Texas Children's Hospital, respectively. Adult blood buffy coats from healthy donors were obtained from the Gulf Coast Regional Blood Center, Texas. For the isolation of thymic CD11c-positive dendritic cells, thymuses were cut into small pieces and digested as previously described³⁰. After separation of the mononuclear cells by Ficoll centrifugation, lineage-negative cells were obtained from thymuses and peripheral blood by negative depletion, and CD11c⁺ lineage[−] $CD4^+$ cells purified using fluorescence-activated cell sorting on a FACSAria (BD Biosciences) to reach >99% purity, as described^{12,14}. Isolated dendritic cells were cultured with various stimuli including human TSLP (15 ng ml^{−1}; recombinant protein was prepared in house by using an adenovirus vector system) as described¹². After 24 h of culture, we determined the cell-surface marker characteristics of the activated dendritic cells, also as described¹².

Isolation of T cells. For the isolation of T-cell lineage thymocytes, thymuses were digested and mononuclear cells were separated by Ficoll centrifugation. T-cell lineage thymocytes were obtained by negative depletion using a mixture of mouse monoclonal antibodies against the lineage markers CD11c, CD14, CD15, CD20, CD56 and CD235a. This was followed by incubation with goat anti-mouse IgG-coated magnetic beads (M-450, Dynal). $CD4^+CD8^-$, $CD4^+CD8^+$, $CD4^+CD8^-$, $CD4^+CD8^+$, $CD4^+CD8^-CD25^+$ and $CD4^+CD8^-CD25^-$ thymocytes were isolated by magnetic bead cell sorting to reach >98.5% purity. $CD4^+CD45RA^+CD25^-$ naive T cells (purity >99%) were isolated by a $CD4^+$ T-cell Isolation Kit II (Miltenyi Biotec) followed by cell sorting. The labelling of T cells with CFSE (Molecular Probes) was performed as described¹². In some experiments, isolated T cells were cultured with plate-bound anti-CD3 antibody (UCHT1; 10 μ g ml^{−1}), soluble anti-CD28 (28.2; 2 μ g ml^{−1}) and 20 ng ml^{−1} of IL-2 and IL-7 (R&D Systems).

Dendritic cell and T cell co-culture. After 24 h of culture, CD11c-positive dendritic cells were collected and washed three times to remove any cytokine. Viable dendritic cells were counted by Trypan-blue exclusion of the dead cells. Remaining cells were co-cultured with 2×10^4 freshly isolated thymocytes or peripheral blood naive $CD4^+$ T cells in round-bottomed 96-well culture plates in Yssel's medium containing 1% human AB serum and 20 ng ml^{−1} IL-7. Cultures were done in triplicate at a 1:2 dendritic cell:T-cell ratio or at increasing ratios. Neutralizing anti-HLA-DR (BD Biosciences), anti-IL-2, anti-CD80 and anti-CD86 (R&D Systems) monoclonal antibodies were used in culture at a concentration of 1×10^{-1} to 1×10^2 μ g ml^{−1} (Fig. 3d) and 30 μ g ml^{−1} (Supplementary Fig. S6).

T-cell proliferation and expansion assay. After 7 days of culture, cells were collected and resuspended in an EDTA-containing medium to dissociate the clusters. Viable cells were counted by Trypan-blue exclusion of the dead cells. The remaining cells were stained with phycoerythrin-conjugated anti-CD25 (M-A251) and allophycocyanin-conjugated anti-CD4 (RPA-T4) antibodies, with or without fluorescein isothiocyanate-conjugated anti-CD8 (SK1) antibody, and analysed with a FACSCalibur (BD Biosciences). Dead cells were excluded on the basis of side- and forward-scatter characteristics, and viable $CD4^+CD25^+$ T-cell numbers were calculated as follows: (the percentage of cells in the cell type) $\times$ (the number of viable cells).

Suppressive function assay. After 7 days of culture, $CD4^+CD25^+$ and

CD4⁺CD25⁻ thymocytes expanded by exposure to anti-CD3 antibody, anti-CD28 antibody, IL-2 and IL-7, and CD4⁺CD25⁺ T cells generated by culturing with TSLP-DCs, were isolated by cell sorting. These three cell types and their mixtures at a 1:1 ratio were then re-stimulated for 5 days by culturing in the presence of 5 µg ml⁻¹ anti-CD3 (HIT3a) antibody, 2 µg ml⁻¹ anti-CD28 (28.2) antibody and 20 ng ml⁻¹ IL-7, after which time the cellular proliferation was assessed by [³H]thymidine incorporation, as described¹².

Dendritic cell and T-cell cytokine production. Dendritic cell culture supernatants were collected at 24 h, and cytokine production was assessed as described¹⁴. CD4⁺CD25⁻ cells from adult peripheral blood or fetal thymuses were cultured for 7 days with TSLP-DCs and IL-7, or with anti-CD3 antibody, anti-CD28 antibody, IL-2 and IL-7. T cells were then washed and re-stimulated with phorbol myristate acetate and ionomycin. Intracellular cytokine staining was performed as described¹².

Reverse transcription-polymerase chain reaction. For conventional reverse transcription-polymerase chain reaction (RT-PCR), the PCR reactions were performed with an initial denaturation step of 95 °C for 5 min, followed by 31 cycles of 94 °C for 1 min, 60 °C for 30 s, 72 °C for 1 min and a final elongation step of 72 °C for 7 min. The real-time quantitative reactions were performed as described¹². Values are expressed as arbitrary units (relative to glyceraldehyde-3-phosphate dehydrogenase (*Gapdh*) levels × 10³). The following primers were used: *Gapdh*: 5'-CCACATCGCTCAGACACCAT-3' and 5'-GGCAACAATATCCACTTTACCAGAGT-3'; *Foxp3*: 5'-GAAACAGCACATTCAGAGTTC-3' and 5'-ATGGCCAGCGGATGAG-3'.

Immunohistochemistry. Immunohistological staining was performed by using anti-human antibodies; mouse anti-CD11c (B-ly6; S-HCL-3), CD25 (M-A251) CD86 (37301), CTLA4 (BNI-3) and DC-LAMP (104.G4) antibodies; and sheep anti-TSLP antibody (R&D Systems), as described¹⁴. In blocking experiments, 2 µg ml⁻¹ of anti-TSLP antibodies were incubated with 50 ng ml⁻¹ of TSLP or IL-7, and used for the staining.

Received 1 April; accepted 6 June 2005.

- Hassall, A. H. *The Microscopic Anatomy of The Human Body in Health and Disease* (Samuel Highly, London, 1849).
- Farr, A. G., Dooley, J. L. & Erickson, M. Organization of thymic medullary epithelial heterogeneity: implications for mechanisms of epithelial differentiation. *Immunol. Rev.* **189**, 20–27 (2002).
- Blau, J. N. A phagocytic function of Hassall's corpuscles. *Nature* **208**, 564–567 (1965).
- Blau, J. N. & Veall, N. The uptake and localization of proteins, Evans Blue and carbon black in the normal and pathological thymus of the guinea-pig. *Immunology* **12**, 363–372 (1967).
- Senelar, R., Escola, M. J., Escola, R., Serrou, B. & Serre, A. Relationship between Hassall's corpuscles and thymocytes fate in guinea-pig foetus. *Biomedicine* **24**, 112–122 (1976).
- Nishio, H., Matsui, K., Tsuji, H., Tamura, A. & Suzuki, K. Immunolocalization of the mitogen-activated protein kinase signalling pathway in Hassall's corpuscles of the human thymus. *Acta Histochem.* **103**, 89–98 (2001).
- He, W., Zhang, Y., Deng, Y. & Kabelitz, D. Induction of TCR-γδ expression on triple-negative (CD3⁻CD4⁻CD8⁻) human thymocytes. Comparative analysis of the effects of IL-4 and IL-7. *J. Immunol.* **154**, 3726–3731 (1995).
- Le, P. T., Lazorick, S., Whichard, L. P., Haynes, B. F. & Singer, K. H. Regulation of cytokine production in the human thymus: epidermal growth factor and transforming growth factor alpha regulate mRNA levels of interleukin 1α (IL-1α), IL-1β, and IL-6 in human thymic epithelial cells at a post-transcriptional level. *J. Exp. Med.* **174**, 1147–1157 (1991).
- Romagnani, P. et al. High CD30 ligand expression by epithelial cells and Hassall's corpuscles in the medulla of human thymus. *Blood* **91**, 3323–3332 (1998).
- Zaitseva, M. et al. Stromal-derived factor 1 expression in the human thymus. *J. Immunol.* **168**, 2609–2617 (2002).
- Annuziato, F. et al. Macrophage-derived chemokine and EBI1-ligand chemokine attract human thymocytes in different stage of development and are produced by distinct subsets of medullary epithelial cells: possible implications for negative selection. *J. Immunol.* **165**, 238–246 (2000).
- Watanabe, N. et al. Human thymic stromal lymphopoietin promotes dendritic cell-mediated CD4⁺ T cell homeostatic expansion. *Nature Immunol.* **5**, 426–434 (2004).
- Leonard, W. J. TSLP: finally in the limelight. *Nature Immunol.* **3**, 605–607 (2002).
- Soumelis, V. et al. Human epithelial cells trigger dendritic cell-mediated allergic inflammation by producing TSLP. *Nature Immunol.* **3**, 673–680 (2002).
- Friend, S. L. et al. A thymic stromal cell line supports in vitro development of surface IgM⁺ B cells and produces a novel growth factor affecting B and T lineage cells. *Exp. Hematol.* **22**, 321–328 (1994).
- Bender, J., Mitchell, T., Kappler, J. & Marrack, P. CD4⁺ T cell division in irradiated mice requires peptides distinct from those responsible for thymic selection. *J. Exp. Med.* **190**, 367–374 (1999).
- Ernst, B., Lee, D. S., Chang, J. M., Sprent, J. & Surh, C. D. The peptide ligands mediating positive selection in the thymus control T cell survival and homeostatic proliferation in the periphery. *Immunity* **11**, 173–181 (1999).
- Hsieh, C.-S. et al. Recognition of the peripheral self by naturally arising CD4⁺CD25⁺ T cell receptors. *Immunity* **21**, 267–277 (2004).
- Salomon, B. et al. B7/CD28 costimulation is essential for the homeostasis of the CD4⁺CD25⁺ immunoregulatory T cells that control autoimmune diabetes. *Immunity* **12**, 431–440 (2000).
- Tai, X., Cowan, M., Feigenbaum, L. & Singer, A. CD28 costimulation of developing thymocytes induces *Foxp3* expression and regulatory T cell differentiation independently of interleukin 2. *Nature Immunol.* **6**, 152–162 (2005).
- Tarbell, K. V., Yamazaki, S., Olson, K., Toy, P. & Steinman, R. M. CD25⁺ CD4⁺ T cells, expanded with dendritic cells presenting a single autoantigenic peptide, suppress autoimmune diabetes. *J. Exp. Med.* **199**, 1467–1477 (2004).
- Stephens, L. A., Mottet, C., Mason, D. & Powrie, F. Human CD4⁺CD25⁺ thymocytes and peripheral T cells have immune suppressive activity in vitro. *Eur. J. Immunol.* **31**, 1247–1254 (2001).
- Sakaguchi, S. Naturally arising CD4⁺ regulatory T cells for immunologic self-tolerance and negative control of immune responses. *Annu. Rev. Immunol.* **22**, 531–562 (2004).
- Fontenot, J. D. & Rudensky, A. Y. Molecular aspects of regulatory T cell development. *Semin. Immunol.* **16**, 73–80 (2004).
- Jordan, M. S. et al. Thymic selection of CD4⁺CD25⁺ regulatory T cells induced by an agonist self-peptide. *Nature Immunol.* **2**, 301–306 (2001).
- Apostolou, I., Sarukhan, A., Klein, L. & von Boehmer, H. Origin of regulatory T cells with known specificity for antigen. *Nature Immunol.* **3**, 756–763 (2002).
- Malek, T. R., Yu, A., Vincek, V., Scibelli, P. & Kong, L. CD4 regulatory T cells prevent lethal autoimmunity in IL-2Rβ-deficient mice: Implications for the nonredundant function of IL-2. *Immunity* **17**, 167–178 (2002).
- Kyewski, B. & Klein, L. Self-antigen presentation by thymic stromal cells: a subtle division of labour. *Curr. Opin. Immunol.* **12**, 179–186 (2000).
- von Boehmer, H. Selection of the T-cell repertoire: receptor-controlled checkpoints in T-cell development. *Adv. Immunol.* **84**, 201–238 (2004).
- Bendris-Vermare, N. et al. Human thymus contains IFN-α-producing CD11c⁺, myeloid CD11c⁺, and mature interdigitating dendritic cells. *J. Clin. Invest.* **107**, 835–844 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. J. Finegold and K. Sternberg for tissue materials; E. Weider, Z. He and K. Ramirez for performing the cell sorting; B. Nortzon and M. Hajek for assistance in the preparation of the manuscript; and L. L. Lanier, S. Ullrich, M. F. Wilkinson and K. Schluns for critical reading and suggestions.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to Y.-J.L. (yjliu@mdanderson.org).

LETTERS

The DNA damage pathway regulates innate immune system ligands of the NKG2D receptor

Stephan Gasser¹, Sandra Orsulic², Eric J. Brown³ & David H. Raulet¹

Some stimulatory receptors of the innate immune system, such as the NKG2D receptor (also called KLRK1) expressed by natural killer cells and activated CD8⁺T cells, recognize self-molecules that are upregulated in diseased cells by poorly understood mechanisms¹. Here we show that mouse and human NKG2D ligands are upregulated in non-tumour cell lines by genotoxic stress and stalled DNA replication, conditions known to activate a major DNA damage checkpoint pathway initiated by ATM (ataxia telangiectasia, mutated) or ATR (ATM- and Rad3-related) protein kinases². Ligand upregulation was prevented by pharmacological or genetic inhibition of ATR, ATM or Chk1 (a downstream transducer kinase in the pathway). Furthermore, constitutive ligand expression by a tumour cell line was inhibited by targeting short interfering RNA to ATM, suggesting that ligand expression in established tumour cells, which often harbour genomic irregularities, may be due to chronic activation of the DNA damage response pathway. Thus, the DNA damage response, previously shown to arrest the cell cycle and enhance DNA repair functions, or to trigger apoptosis, may also participate in alerting the immune system to the presence of potentially dangerous cells.

To investigate mechanisms leading to NKG2D ligand upregulation, we examined two transformed ovarian epithelial cell lines from *p53*^{-/-} mice³. The C1 cell line had been transduced with *K-ras* and *c-myc*, whereas the C2 cell line had been transduced with *Akt* and *c-myc* (Fig. 1a). Both transformed cell lines grew well in cell culture but did not express appreciable levels of mouse NKG2D ligands, as detected by staining with a tetrameric NKG2D reagent that binds to all mouse NKG2D ligands (Rae1, MULT1 and H60 (ref. 1)) (Fig. 1b). Ligand upregulation failed to occur when C1 and C2 cells were transfected or super-transduced with numerous other oncogenes, including *E6*, *E7*, *E1A* or *Ras* V12 (data not shown), some of which interfere with expression of the retinoblastoma tumour suppressor gene. When injected into the ovaries of nude mice, both cell lines generated ovarian epithelial tumours, which were established as cell lines T1 and T2 (ref. 3). Both T1 and T2 exhibited significant upregulation of NKG2D ligands (Fig. 1B), including Rae1 (see below). These findings suggested that ligand upregulation was not associated with transformation *per se*.

Ligand expression by C1 or C2 cells was not upregulated by numerous cell stress conditions, including heat shock, hyperoxia, hypoxia, inhibition of the cell cycle (by roscovitine), exposure to inflammatory cytokines such as tumour necrosis factor, interferon or interleukin (IL)-6, or incubation in medium of pH 6 or pH 8.5 (Fig. 2a, Supplementary Fig. S1 and data not shown). In contrast, NKG2D ligands were upregulated in C1 or C2 cells exposed to high doses of ionizing radiation, inhibitors of DNA replication such as mitomycin C, hydroxyurea, 5-fluorouracil (5-FU) and the

DNA polymerase inhibitor aphidicolin, or chromatin-modifying treatments such as trichostatin A, chloroquine and hypotonic conditions⁴ (Fig. 2b; see also Supplementary Fig. S1). These same treatments induced ligand upregulation in cultures of adult fibroblasts, showing that neither transformation nor *p53* deficiency are essential for ligand upregulation (Fig. 2b). In the case of fibroblasts, but less so with the C1 and C2 cell lines, ligand upregulation was also induced by other DNA damaging conditions such as ultraviolet light

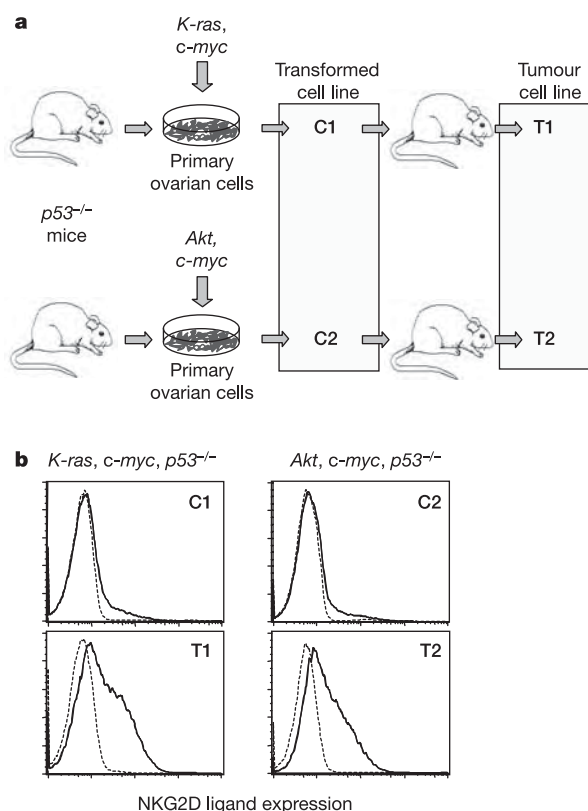


Figure 1 | NKG2D ligand upregulation is associated with tumorigenesis, not with transformation *per se*. **a**, Cell lines from transgenic mice expressing the avian retrovirus receptor. Ovarian epithelial cells from *p53*^{-/-} mice were transduced with the indicated genes and established as cell lines C1 and C2. Implantation of C1 and C2 cells into ovaries of nude mice led to formation of tumours, which were established as cell lines T1 and T2. **b**, Transformed ovarian epithelial cell lines C1 and C2, and T1 and T2 tumour cell lines, were stained with NKG2D tetramers (solid line) or with control tetramers (dashed line).

¹Department of Molecular and Cell Biology and Cancer Research Laboratory, University of California, Berkeley, California 94720-3200, USA. ²Massachusetts General Hospital Center for Cancer Research, Harvard Medical School, Charlestown, Massachusetts 02129, USA. ³University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104-6160, USA.

and the chemotherapy agents cisplatin and ara-C (Fig. 2b; see also Supplementary Fig. S1). Cell surface ligand upregulation detected with NKG2D tetramers or Rae1 antibodies (Fig. 2b) was accompanied by increases in both Rae1 and MULT-1 messenger RNAs (Fig. 3a; see also Supplementary Fig. S2). Both the mRNA and protein levels were initially detected at 3–5 h, and reached a plateau after 16–24 h (Fig. 3a, b). Ligand levels reverted to background levels 72 h after aphidicolin was washed out of the medium (data not shown).

Similarly, the homologous human NKG2D ligands ULBP1, ULBP2 and ULBP3 were often upregulated in secondary human foreskin fibroblasts treated with high doses of ionizing radiation and inhibitors of DNA replication, including aphidicolin, but not with roscovitine (Fig. 2c). Mitomycin C and aphidicolin also modestly upregulated MICA, another human NKG2D ligand (Fig. 2c and Supplementary Fig. S1).

Mouse T cell blasts treated with aphidicolin also upregulated NKG2D ligands modestly, and exhibited greater sensitivity to lysis by IL-2-activated natural killer (NK) cells (Fig. 2d). Lysis was inhibited by NKG2D antibody, indicating that upregulated ligands induce elevated lysis, but the partial effect suggests that aphidicolin may also induce other NK target ligands.

The treatments that induced ligand upregulation all activate a major DNA damage response pathway initiated by ATR and/or ATM (depending on the treatment). The pathway is mediated by downstream mediators including the Chk1 and Chk2 kinases and p53 (refs 2,5,6), and results in cell cycle arrest, activation of DNA repair pathways, new transcription, and, if damage is extensive, apoptosis. Consistent with a role for this pathway in NKG2D ligand upregulation, phosphorylation of Chk1 on serine 345, which is required for activation of this kinase, occurred before ligand upregulation was detectable within 1 h of treatment of fibroblasts with aphidicolin (Fig. 3c). Phosphorylation of Chk2 on threonine 387 also occurred but was delayed.

The role of ATR in aphidicolin-induced NKG2D ligand expression was investigated with three independent approaches. Ligand upregulation in response to aphidicolin was blocked by caffeine, an inhibitor

of ATR and ATM⁷, at doses close to the IC₅₀ (half-maximal inhibitory concentration) for ATR (1.1 mM, Fig. 4a). In a second approach, ATR gene expression was specifically impaired using siRNAs in adult fibroblast cultures, introduced by transduction with green fluorescent protein (GFP)-expressing retroviruses. Upregulation of NKG2D ligands induced by aphidicolin was strongly inhibited in cells transduced with *Atr* siRNA compared to untransduced (GFP⁺) cells in the same cultures or to cells transduced with control siRNA (Fig. 4b). Cells infected with the control siRNA or *Atr* siRNA but not treated with aphidicolin did not upregulate ligands (Fig. 4b). As a third approach, the ATR gene was deleted in cultured cells. ATR function is required for cell viability, necessitating the use of a conditional *Atr* mutation. A fibroblast cell line was derived from adult mice with a gene-targeted *Atr* allele in which two critical exons were flanked by *loxP* sites⁸. Fibroblasts transduced with a Cre recombinase-expressing retrovirus to inactivate the *Atr* gene and treated with aphidicolin showed minimal upregulation of NKG2D ligand cell surface proteins (Fig. 4c), or mRNAs (Fig. 4d), compared with cells transduced with a control murine stem cell virus (MSCV) retrovirus. Control experiments established that both the untreated Cre-transduced *Atr*^{fllox} cells and the untreated MSCV-transduced *Atr*^{fllox} cells proliferated actively (Supplementary Fig. S3), a necessary condition for activation of the replication checkpoint. Proliferation of both cell populations was blocked with aphidicolin (Supplementary Fig. S3). Cre transduction had no effect on aphidicolin-treated wild-type fibroblasts (Fig. 4c, d). Thus, upregulation of NKG2D ligands as a result of replication arrest is *Atr*-dependent.

A role for the Chk1 kinase was suggested by the finding that aphidicolin-induced ligand upregulation was inhibited by three independent Chk1 inhibitors (Fig. 5a). The effective doses were close to the IC₅₀ values of all three inhibitors for Chk1 (8 nM staurosporine, 15 nM SB-218078 and 3 μ M debromohymenialdisine (DBH)) (Fig. 5a). Each of these inhibitors also acts on targets other than Chk1, but the IC₅₀ values differ and the other targets are not all shared by the three inhibitors. As a more specific test, transduction of fibroblasts with *Chk1* siRNA reproducibly inhibited aphidicolin-induced NKG2D ligand upregulation as compared to untransduced

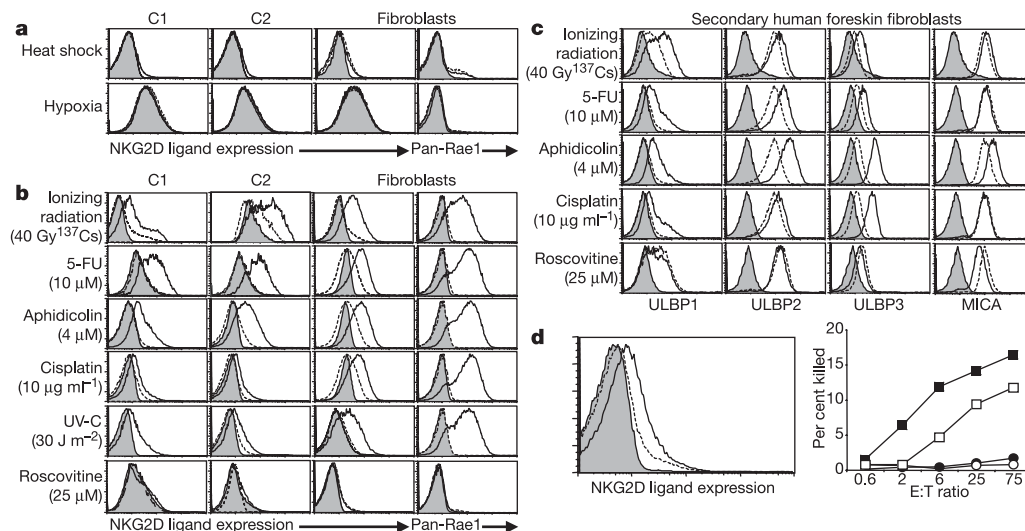


Figure 2 | NKG2D ligands are induced by DNA-damaging agents and DNA synthesis inhibitors. **a**, No induction of ligands in C1 or C2 cells by heat shock (42 °C, 90 min) and hypoxia (1% O₂, 48 h). Cells were stained with NKG2D tetramers or Rae1 antibody. **b**, **c**, Upregulation of ligands in mouse cells (**b**) or human fibroblasts (passage 3–6) (**c**) by DNA-damaging agents and DNA synthesis inhibitors (16 h treatment). Treated cells (solid lines) were compared to untreated cells (dashed lines), or to treated cells stained with control tetramers or isotype control antibodies (filled histograms). **d**, Aphidicolin treatment upregulates NKG2D ligands on T-cell blasts

(left panel, key as defined above) and elevates their sensitivity to IL-2-activated NK cells (right). NK cells were incubated with treated (squares) or untreated (circles) target cells in the absence (filled) or presence (open) of NKG2D antibody. The effector to target ratio (E:T) is indicated. In repetitions of the experiment, untreated cells were killed but always substantially less well than treated cells. For all the flow cytometry histograms in **a–d**, the x axis depicts staining intensity and the y axis depicts the relative number of cells.

(GFP⁻) cells in the same culture or cells transduced with control siRNA (Fig. 5b). The *Chk1* siRNA did not completely abrogate Chk1 protein expression (Fig. 5c), possibly explaining why pharmacological inhibitors were more effective at blocking ligand upregulation. The inhibitor and siRNA results indicate that Chk1 is necessary for optimal NKG2D ligand upregulation in response to aphidicolin treatment.

Ionizing radiation, hypotonic conditions and other treatments that activate ATM, in some cases preferentially to ATR^{4,9}, also induced upregulation of NKG2D ligands (Fig. 2; see also Supplementary Fig. S1). An ATM siRNA/GFP retrovirus partially inhibited NKG2D ligand upregulation in cells exposed to ionizing radiation or hypotonic conditions, but had no effect on ligand upregulation in cells exposed to aphidicolin or ultraviolet C, which preferentially activate ATR⁹ (Supplementary Fig. S4). Ligand upregulation in response to ATM inducers was also inhibited by caffeine, which blocks both ATR and ATM⁷ (Supplementary Fig. S4a). Thus, induction of NKG2D ligands requires ATM or ATR, depending on the nature of the treatment.

It is unlikely that NKG2D ligand upregulation is an indirect consequence of ATR/ATM-induced cell cycle arrest, because cell cycle arrest induced by the cyclin-dependent kinase inhibitor roscovitine or serum deprivation did not trigger ligand upregulation (Fig. 2b, c; see also Supplementary Fig. S1). It is also unlikely that ligand upregulation is a property of cells entering the cell death

pathway, because such cells, defined by staining with Annexin-V, did not upregulate NKG2D ligands significantly (data not shown). Hence, the DNA damage response probably has a more direct role in the process of ligand upregulation.

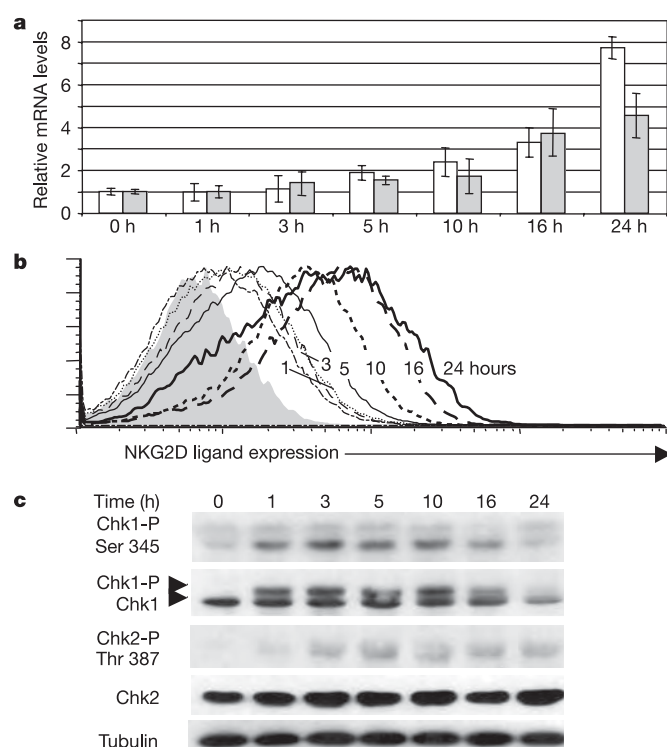


Figure 3 | Kinetics of the upregulation of NKG2D ligands and phosphorylation of the Chk1 kinase. **a**, Levels of Rael1 (open bar) and MULT1 (filled bar) transcripts determined by real-time RT-PCR in cultured fibroblasts treated with aphidicolin for the indicated times. The Rael1 primers amplify all Rael1 isoforms. Data represent means $\pm$ s.d., $n = 3$ independent experiments. **b**, Cell surface expression of NKG2D ligands on cells from **a**. Cells treated for the indicated period were compared to untreated cells (dashed and dotted lines), or to 24 h-treated cells stained with control tetramers (filled histograms). The y axis depicts the relative number of cells. **c**, Kinetics of phosphorylation of Chk1 in aphidicolin-treated fibroblasts. Phosphorylation was detected by western blotting as a shift in the mobility of Chk1, or with the use of a phospho-Ser 345 Chk1 specific antibody or phospho-Thr 387 Chk2 specific antibody.

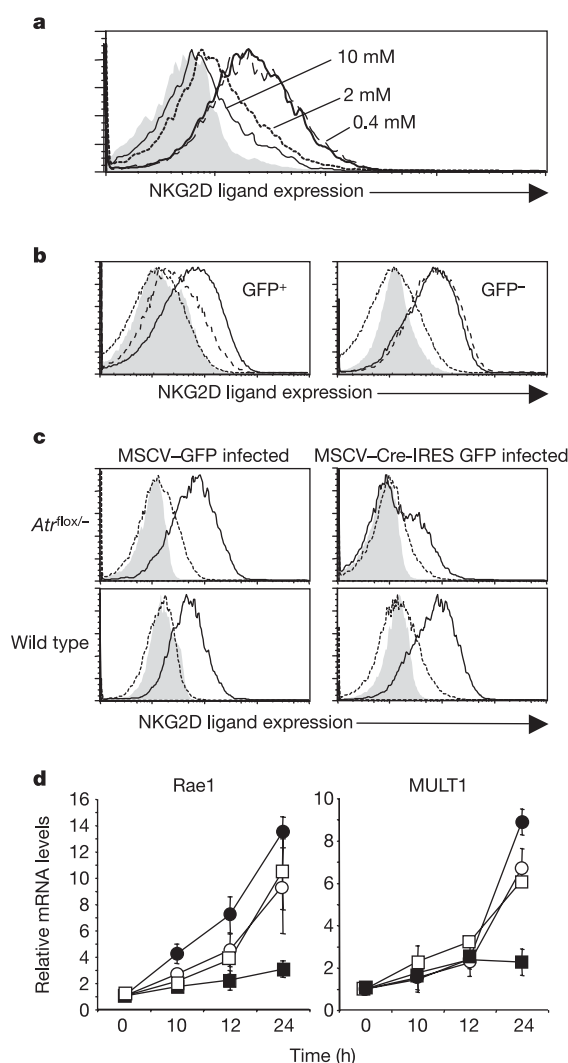


Figure 4 | Ligand upregulation in aphidicolin-treated fibroblasts is dependent on ATR function. **a**, Ligand upregulation inhibited by caffeine (an inhibitor of ATR/ATM) added 1 h before aphidicolin. Aphidicolin-treated cells incubated with the indicated caffeine concentrations were compared to treated cells incubated with no inhibitor (thick, solid line), or to treated cells stained with control tetramers (filled histograms). **b**, Inhibition of ligand upregulation by *Atr* siRNA. Fibroblasts were transduced with retroviral vectors encoding GFP and *Atr* siRNA (dashed lines in histograms) or GFP and control siRNA (solid lines), cultured for 5 days and treated with 4 μ M aphidicolin for 16 h. NKG2D ligand expression was determined by gating on transduced (GFP⁺) and untransduced (GFP⁻) cells from the same cultures. The filled histogram indicates aphidicolin-treated cells stained with NKG2D tetramers; the dotted line indicates untreated cells stained with NKG2D tetramers. **c**, Conditional deletion of the ATR gene prevents ligand upregulation. *Atr*^{flox/-} or wild-type fibroblasts were transduced with MSCV-IRES-GFP or MSCV-Cre-IRES-GFP, cultured and analysed as in **b**, gating on transduced (GFP⁺) cells. The solid line indicates treated cells stained with NKG2D tetramers; the filled histogram indicates treated cells stained with control tetramers; and the dotted line indicates untreated cells stained with NKG2D tetramers. **d**, Reduced induction of Rael1 and MULT1 transcripts in cells deficient for ATR expression. Real-time PCR of cDNA samples from transduced fibroblast cultures described in **c**. Open symbols, wild type fibroblasts; filled symbols, *Atr*^{flox/-} fibroblasts; circles, control; squares, Cre-transduced. Means $\pm$ s.d. of three independent experiments are shown.

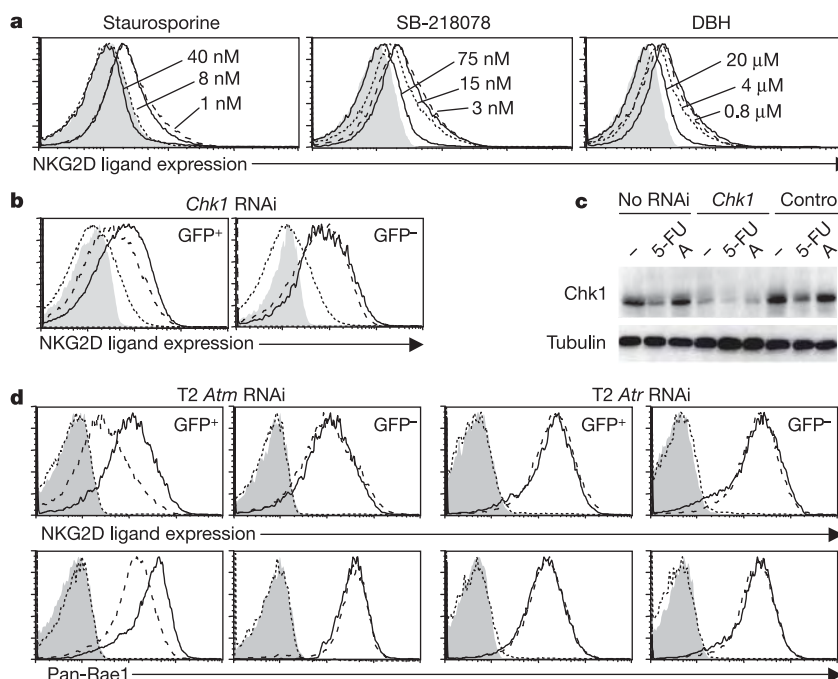


Figure 5 | Role of Chk1 in expression of NKG2D ligands in aphidicolin-treated fibroblasts and tumour cell lines. **a**, Chk1 inhibitors prevent upregulation of NKG2D ligands. Fibroblasts were pre-incubated with staurosporine, SB-218078, debromohymenialdisine, or DMSO before aphidicolin treatment. The thick, solid line represents treated cells (no inhibitor) stained with NKG2D tetramers; filled histogram represents treated cells (no inhibitor) stained with control tetramers **b**, Inhibition of ligand upregulation by *Chk1* siRNA. Fibroblasts were transduced with retroviral vectors encoding GFP and *Chk1* siRNA (dashed lines in histograms) or GFP and control siRNA (solid lines) before treatment, as in Fig. 4b. The filled histogram represents aphidicolin-treated cells stained with control tetramers; the dotted line indicates untreated cells stained with NKG2D tetramers. **c**, Chk1 protein levels are reduced in cells transduced

with *Chk1* siRNA. Untransduced or sorted GFP⁺ transduced fibroblasts were cultured for 2 days and treated for 16 h with DMSO (labelled –), 4 μ M aphidicolin (labelled A) or 10 μ M 5-FU. Lysates were prepared and western blotted with Chk1 antibodies. **d**, *Atm* siRNA, but not *Atr* siRNA, decreases levels of NKG2D ligands in T2 cells. T2 cells were transduced with GFP retroviral vectors encoding control siRNA (solid lines) or either *Atm* or *Atr* siRNA (dashed lines in histograms). After culturing for 7 days, NKG2D ligand expression was determined by gating on transduced (GFP⁺) and untransduced (GFP[–]) cells from the same cultures. Filled histogram, control siRNA transduced T2 cells stained with control tetramer; dotted line, *Atr* or *Atm* siRNA transduced T2 cells stained with control tetramers. For all the flow cytometry histograms in **a**, **b**, **d**, the x axis depicts staining intensity and the y axis depicts the relative number of cells.

Activation of the DNA damage pathway may represent a more distinctive feature of diseased cells than other correlates of tumorigenesis or infection. The pathway is activated in cells infected with several viruses¹⁰, raising the possibility that it has a role in ligand upregulation after viral infection¹¹. The pathway is also frequently activated in tumour cell lines^{12–14}, possibly due to the greater genomic instability of these cells as compared to transformed cells^{13,14}. Furthermore, recent reports have demonstrated the activation of ATM and other components of the DNA damage response pathway in pre-cancerous lesions, at a stage before genomic instability is apparent^{15,16}. We addressed whether chronic activation of the DNA damage pathway contributes to constitutive expression of NKG2D ligands observed in the T2 ovarian tumour cell line (Fig. 1). T2 cells transduced with *Atm* siRNA exhibited reduced NKG2D ligand expression, whereas *Atr* siRNA had no effect (Fig. 5d). Thus, constitutive ligand expression in tumour cell lines may be dependent on ATM activity, at least in some instances. Accelerated loss of genomic stability or other cellular changes that activate the DNA damage response in tumour cells may lead to constitutive ATM activation. Further studies will be necessary to establish the generality of the link between the DNA damage response and NKG2D ligand expression in tumour cells.

Transcriptional regulators including the p53 tumour suppressor, p73 (ref. 17) and p63 are activated by Chk1. p53 is not required for ligand upregulation, which occurred in the *p53*^{–/–} C1 and C2 cell lines, but p73 or p63 may have redundant or unique roles in ligand upregulation. The p53-independent component of NKG2D ligand upregulation means that loss of *p53* during tumorigenesis should

not by itself fully disable the immune-enhancing potential of the cells.

These findings suggest a novel link between the immune response and processes that regulate genome integrity, and may have clinical significance. It is possible that part of the efficacy of some chemotherapies and radiotherapies, most of which activate the DNA damage response^{18,19}, is due to the induction of NKG2D ligands and consequent enhanced sensitivity of the cell to the immune system. The pathway leading to the upregulation of NKG2D ligands may be a productive target for design of therapeutic agents to enhance the immunogenicity of tumour cells while reducing overall toxicity.

METHODS

Mice, cells and cell treatments. C57BL/6 mice from the Jackson Laboratory were bred and housed as described²⁰. Ear- and tail-derived fibroblasts from adult C57BL/6 mice or 129SvEv/C57BL/6-*Atr*^{lox/+} mice^{8,21} were established as described²². Unless otherwise specified, the ear fibroblasts were used for the experiments. C1, C2, T1 and T2 mouse ovarian cancer cell lines were generated as described³ and treated for 16 h unless stated otherwise. Secondary neonatal human dermal foreskin fibroblasts were purchased (Cascade Biologics) and cultured in 106 medium.

Hypotonic swelling was carried out for 16 h in PBS containing 0.45% glucose (w/v) and 2% FBS, with the NaCl concentration adjusted to 50 mM or, as a control, 140 mM. For inhibitor experiments, inhibitors were added to subconfluent cultures beginning 1 h before addition of 4 μ M aphidicolin, and ligand expression was determined 16 h later.

Reagents. 5-bromo-2'-deoxyuridine (BrdU), 5-fluorouracil (5-FU), aphidicolin, caffeine, *cis*-diammineplatinum(II) dichloride, cytosine β -D-arabino-furanoside

hydrochloride, mitomycin C, hydroxyurea, staurosporine and *t*-butylhydroquinone were purchased from Sigma. Roscovitine, SB-218078 and debromohymenialdisine were purchased from Calbiochem. NKG2D tetramers and control tetramers were produced as previously described²³. For staining, we used Pan-Rae1, MICA, ULBP1, ULBP2 and ULBP3 specific monoclonal antibodies (R&D Systems), mouse IgG and F(ab')₂ fragment goat anti-mouse or rat IgG plus IgM coupled to phycoerythrin (Jackson ImmunoResearch Laboratories).

MSCV-Cre constructs and transduction of fibroblasts. *Cre* was subcloned adjacent to the cytomegalovirus promoter in the pMSCV2.2-IRES-GFP proviral vector (gift of W. Sha). Retroviral supernatants were generated as described²⁴. **siRNA retroviral constructs.** *Chk1* siRNA (5'-CAACTTGCTGTGAATAGAAT-3') corresponded to the mouse counterpart of a published human *Chk1* siRNA²⁵. *Atr* siRNA (5'-AGGAAGCAATTCCACATTAAAGC-3') and *Atm* siRNA (5'-GAG GTGGCTCTTATTCTAC-3') were selected based on Dharmacon's siRNA Design Center algorithm (Dharmacon). The control siRNA (5'-AACAAGTGAAGCAG TCGCAGT-3') has been described previously²⁶. The siRNAs were subcloned in the pMND-Banshee vector (a gift of J. Alberola-Ila), as described previously²⁷. The control RNAi-pMND-Banshee, *ATR*-pMND-Banshee and *Chk1*-pMND-Banshee plasmids were transiently transfected into 293 T cells using Lipofectamine 2000 according to the manufacturer's instructions (Invitrogen). 48 h after transfection, 50% confluent cultures of fibroblasts or T2 were transduced as described²⁴.

As determined by real time RT-PCR in sorted populations, the level of *ATR* transcripts in *Atr*-siRNA-transduced cells in Fig. 4b was $37 \pm 3\%$ of the level in untransduced (GFP⁻) cells in the same culture (mean $\pm$ s.d., $n = 3$). In Fig. 5d, *Atm* mRNA in *Atm*-siRNA-transduced T2 cells was $53 \pm 2\%$ of the level in untransduced (GFP⁻) cells in the same culture, whereas *Atr* mRNA in *Atr*-siRNA-transduced cells was $49 \pm 5\%$ of the level in untransduced (GFP⁻) cells in the same culture. Because of impurities in the sorted populations, the knockdowns may be more efficient than indicated.

Western blotting. Whole-cell extracts were prepared from untreated or treated populations, electrophoresed in 6% or 8% SDS-PAGE gels, and blotted onto nitrocellulose membranes. Antibodies against Chk1 (sc-8408, Santa Cruz Biotechnology), phospho-Chk1-Ser 345, Chk2, phospho-Chk2-Thr 387 (all Cell Signalling Technology), tubulin (CP06, Calbiochem) and horseradish peroxidase-coupled second stage reagents were used to develop the blots (SuperSignal West Pico Stable Solution). Blots were exposed on X-ray film.

Quantitative real-time RT-PCR. Total RNA was isolated using the RNeasy kit (Qiagen). Real-time PCR assays were performed using an Applied Biosystems 5700 sequence detector. Two micrograms of total RNA was reverse transcribed with random hexamers using a Transcriptor First Strand cDNA Synthesis Kit (Roche). Each amplification mixture (25 μ l) contained 25 ng of reverse-transcribed RNA, 8 μ M forward primer, 8 μ M reverse primer and 12.5 μ l of iTaq SYBR Green Supermix with ROX (Bio-Rad). PCR thermocycling parameters were 50 °C for 2 min, 95 °C for 10 min, and 40 cycles of 95 °C for 15 s, 60 °C for 15 s and 72 °C for 1 min. All samples were normalized to the signal generated from the housekeeping genes *β -actin* or *Gapdh* (for *ATR* and *ATM* levels). SYBR green PCR was performed in triplicate. The following primers were used: *β -actin*-5', 5'-TGTTTGAGACCTTCAACACC-3'; *β -actin*-3', 5'-TAGGAGCC AGAGCAGTAATC-3'; *GAPDH*-5', 5'-GAAGGTCGGTGTGAACCGA-3'; *GAPDH*-3', 5'-GTTAGT GGGGTCTCGTCTCT-3'; *MULTI-1*-5', 5'-CAATGT CTCTGTCCTCGGA-3'; *MULTI-1*-3', 5'-CTGAACACGCTCAGGCACCT-3'; *Pan-Rae1*-5', 5'-TGGACACTCACAAGACCAATG-3'; *Pan-Rae1*-3', 5'-CCCA GGTGGCACTAGGAGT-3'; *ATR*-5', 5'-TGCGCTCTGCTAGAGCAGGT-3'; *ATR*-3', 5'-AGTGCTGGCTGGCTGTGCTG-3'; *ATM*-5', 5'-ATCCAGGCCCTG CAGAATTGGG-3'; *ATM*-3', 5'-CTCCACGCCGCTGTGAACG-3'. Samples prepared without reverse transcription served as negative control templates.

Cytolysis assay. BALB/c spleen cells were activated for 3 days with 2.5 μ g ml⁻¹ concanavalin A, washed with 50 mM α -methyl mannose and recultured with 10 U ml⁻¹ IL-2 plus either aphidicolin or DMSO for 18 h before labelling with ⁵¹Cr for use as target cells. NK cells were prepared as described²⁰. Thirty minutes before adding target cells to initiate the 3-h cytolysis assay²⁰, antibody MI-6 (anti-NKG2D²⁰) was added to a concentration of 50 μ g ml⁻¹ in some groups. Spontaneous release was 5% for DMSO-treated target cells and 15% for aphidicolin-treated target cells.

Received 15 April; accepted 6 June 2005.

Published online 3 July 2005.

1. Raulet, D. H. Roles of the NKG2D immunoreceptor and its ligands. *Nature Rev. Immunol.* **3**, 781–790 (2003).

2. Abraham, R. T. Cell cycle checkpoint signalling through the ATM and ATR kinases. *Genes Dev.* **15**, 2177–2196 (2001).
3. Xing, D. & Orsulic, S. A genetically defined mouse ovarian carcinoma model for the molecular characterization of pathway-targeted therapy and tumour resistance. *Proc. Natl Acad. Sci. USA* **102**, 6936–6941 (2005).
4. Bakkenist, C. J. & Kastan, M. B. DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. *Nature* **421**, 499–506 (2003).
5. Osborn, A. J., Elledge, S. J. & Zou, L. Checking on the fork: the DNA-replication stress-response pathway. *Trends Cell Biol.* **12**, 509–516 (2002).
6. Sancar, A., Lindsey-Boltz, L. A., Unsal-Kacmaz, K. & Linn, S. Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annu. Rev. Biochem.* **73**, 39–85 (2004).
7. Sarkaria, J. N. *et al.* Inhibition of ATM and ATR kinase activities by the radiosensitizing agent, caffeine. *Cancer Res.* **59**, 4375–4382 (1999).
8. Brown, E. J. & Baltimore, D. Essential and dispensable roles of ATR in cell cycle arrest and genome maintenance. *Genes Dev.* **17**, 615–628 (2003).
9. Nyberg, K. A., Michelson, R. J., Putnam, C. W. & Weinert, T. A. Toward maintaining the genome: DNA damage and replication checkpoints. *Annu. Rev. Genet.* **36**, 617–656 (2002).
10. Weitzman, M. D., Carson, C. T., Schwartz, R. A. & Lilley, C. E. Interactions of viruses with the cellular DNA repair machinery. *DNA Repair (Amst.)* **3**, 1165–1173 (2004).
11. Lodoen, M. *et al.* NKG2D-mediated natural killer cell protection against cytomegalovirus is impaired by viral gp40 modulation of retinoic acid early inducible 1 gene molecules. *J. Exp. Med.* **197**, 1245–1253 (2003).
12. Bartek, J. & Lukas, J. Chk1 and Chk2 kinases in checkpoint control and cancer. *Cancer Cell* **3**, 421–429 (2003).
13. Mills, K. D., Ferguson, D. O. & Alt, F. W. The role of DNA breaks in genomic instability and tumorigenesis. *Immunol. Rev.* **194**, 77–95 (2003).
14. Sieber, O. M., Heinemann, K. & Tomlinson, I. P. Genomic instability—the engine of tumorigenesis? *Nature Rev. Cancer* **3**, 701–708 (2003).
15. Gorgoulis, V. G. *et al.* Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. *Nature* **434**, 907–913 (2005).
16. Bartkova, J. *et al.* DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. *Nature* **434**, 864–870 (2005).
17. Gonzalez, S., Prives, C. & Cordon-Cardo, C. p73alpha regulation by Chk1 in response to DNA damage. *Mol. Cell Biol.* **23**, 8161–8171 (2003).
18. Bignami, M., Casorelli, I. & Karan, P. Mismatch repair and response to DNA-damaging antitumour therapies. *Eur. J. Cancer* **39**, 2142–2149 (2003).
19. Zhou, B. B. & Bartek, J. Targeting the checkpoint kinases: chemosensitization versus chemoprotection. *Nature Rev. Cancer* **4**, 216–225 (2004).
20. Jamieson, A. M. *et al.* The role of the NKG2D immunoreceptor in immune cell activation and natural killing. *Immunity* **17**, 19–29 (2002).
21. Brown, E. J. & Baltimore, D. ATR disruption leads to chromosomal fragmentation and early embryonic lethality. *Genes Dev.* **14**, 397–402 (2000).
22. Lander, M. R., Moll, B. & Rowe, W. P. A procedure for culture of cells from mouse tail biopsies: brief communication. *J. Natl Cancer Inst.* **60**, 477–478 (1978).
23. Diefenbach, A., Jamieson, A. M., Liu, S. D., Shastri, N. & Raulet, D. H. Ligands for the murine NKG2D receptor: expression by tumour cells and activation of NK cells and macrophages. *Nature Immunol.* **1**, 119–126 (2000).
24. Diefenbach, A., Hsia, J. K., Hsiung, M. Y. & Raulet, D. H. A novel ligand for the NKG2D receptor activates NK cells and macrophages and induces tumour immunity. *Eur. J. Immunol.* **33**, 381–391 (2003).
25. Krause, D. R. *et al.* Suppression of T cell-like kinase activity after DNA damage or replication block requires ATM, NBS1 and Chk1. *Oncogene* **22**, 5927–5937 (2003).
26. Zhang, X. *et al.* Artemis is a phosphorylation target of ATM and ATR and is involved in the G2/M DNA damage checkpoint response. *Mol. Cell Biol.* **24**, 9207–9220 (2004).
27. Hernandez-Hoyos, G., Anderson, M. K., Wang, C., Rothenberg, E. V. & Alberola-Ila, J. GATA-3 expression is controlled by TCR signals and regulates CD4/CD8 differentiation. *Immunity* **19**, 83–94 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Thorner and S. Linn for discussions; Y. Natanzon and L. Zhang for technical assistance; H. Nolla for assistance with flow cytometry; and L. Fritz-Laylin for help with transductions. This work was supported by a National Institutes of Health grant to D.H.R. and an award from the Prostate Cancer Foundation.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.H.R. (raulet@uclink.berkeley.edu).

Cryptic chlorination by a non-haem iron enzyme during cyclopropyl amino acid biosynthesis

Frédéric H. Vaillancourt^{1*}, Ellen Yeh^{1*}, David A. Vosburg^{1†}, Sarah E. O'Connor^{1†} & Christopher T. Walsh¹

Enzymatic incorporation of chlorine, bromine or iodine atoms occurs during the biosynthesis of more than 4,000 natural products¹. Halogenation can have significant consequences for the bioactivity of these products so there is great interest in understanding the biological catalysts that perform these reactions. Enzymes that halogenate unactivated aliphatic groups have not previously been characterized. Here we report the activity of five proteins—CmaA, CmaB, CmaC, CmaD and CmaE—in the construction of coronamic acid (CMA; 1-amino-1-carboxy-2-ethylcyclopropane), a constituent of the phytotoxin coronatine synthesized by the phytopathogenic bacterium *Pseudomonas syringae*². CMA derives from L-*allo*-isoleucine, which is covalently attached to CmaD through the actions of CmaA, a non-ribosomal peptide synthetase module, and CmaE, an unusual acyltransferase. We show that CmaB, a member of the non-haem Fe²⁺, α -ketoglutarate-dependent enzyme superfamily, is the first of its class to show halogenase activity, chlorinating the γ -position of L-*allo*-isoleucine. Another previously undescribed enzyme, CmaC, catalyses the formation of the cyclopropyl ring from the γ -Cl-L-*allo*-isoleucine product of the CmaB reaction. Together, CmaB and CmaC execute γ -halogenation followed by intramolecular γ -elimination, in which biological chlorination is a cryptic strategy for cyclopropyl ring formation.

Coronatine is a hybrid non-ribosomal peptide-polyketide leaf toxin containing CMA, a cyclopropyl amino acid, itself derived from the non-proteinogenic *allo* diastereomer of L-isoleucine^{3,4} (Fig. 1a). A variant of coronatine containing norcoronamic acid (norCMA) derived from L-valine is also produced in low quantities⁵. Although the source of L-*allo*-Ile in *P. syringae* is not known, a cluster of *cma* genes^{2,6} is responsible for generating CMA from L-*allo*-Ile by using non-ribosomal peptide synthetase (NRPS) modules (Fig. 1b). Multimodular NRPS enzymes sequester amino-acid monomers through thioester linkage to the phosphopantetheinyl prosthetic group of carrier-protein domains. Downstream domains or dedicated tailoring enzymes may then catalyse modifications of the tethered monomer or peptide bond formation between monomers. Previously, CmaA, which contains NRPS adenylation (A) and thiolation (T) domains, was shown to activate L-*allo*-Ile as the AMP ester and install it on the 10-kDa T domain⁶. We now assign functions to CmaB, CmaC, CmaD and CmaE and reveal the underlying enzymatic logic by which unactivated carbon centres are functionalized by halogenation to facilitate cyclopropyl formation.

CmaD is a stand-alone 8-kDa T domain that lacks a corresponding A domain but may be aminoacylated *in trans* by the A domain of CmaA (see Supplementary Table S2 for kinetic data of the CmaA A domain). After heterologous expression, apo-CmaD is modified post-translationally with the phosphopantetheine arm to generate

the holo form of the carrier domain. However, CmaA was unable to transfer L-*allo*-[³H]Ile or L-[¹⁴C]Val in the presence of ATP to CmaD, because no radioactivity was associated with the acid-precipitated protein. However, after the addition of CmaE, a 32-kDa protein with homology to the α/β hydrolase superfamily, covalent attachment of these amino acids to CmaD was detected. Autoradiography of the labelled proteins shows time-dependent labelling of CmaA followed by transfer of L-[¹⁴C]Val to CmaD by the action of CmaE (Fig. 2a; see Supplementary Methods for detailed assay protocol). A functional CmaA T domain is required because no labelling of CmaD was observed when the S542A variant of CmaA was used for transfer (Ser 542 is the site of modification for the phosphopantetheine attachment; Fig. 2b). CmaE is a previously undescribed acyltransferase that shuttles amino acid groups between T domains. Because

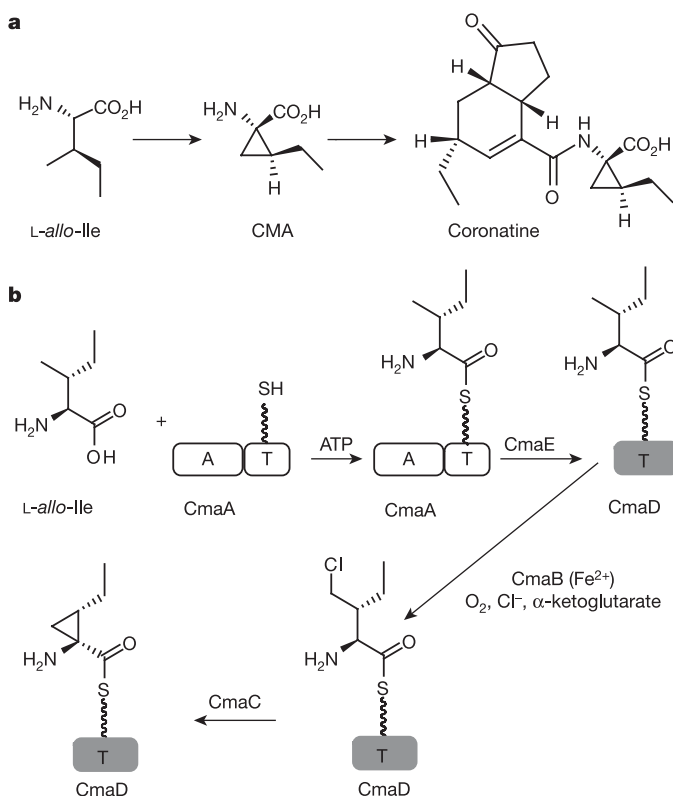


Figure 1 | Biosynthesis of CMA and coronatine. **a**, The biosynthesis of coronatine from L-*allo*-Ile. **b**, The generation of CMA-S-CmaD from L-*allo*-Ile by CmaA, CmaB, CmaC, CmaD and CmaE.

¹Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA. [†]Present address: Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA (S.E.O.); Department of Chemistry, Harvey Mudd College, Claremont, California 91711, USA (D.A.V.).

*These authors contributed equally to this work.

CmaE itself becomes labelled during the reaction, formation of an acyl-enzyme intermediate through the active-site cysteine (Cys 105) is probably involved. Consistent with this was our observation that loading of CmaD *in trans* is abolished when the variant C105A of CmaE is tested for transfer (Fig. 2b). As discussed below, CmaD, but not the T domain of CmaA, is able to present aminoacyl groups for subsequent transformation by CmaB and CmaC.

A second previously undescribed enzyme in this pathway is CmaB, which has homology to α -ketoglutarate (α -KG)-dependent non-haem Fe^{2+} enzymes that use one Asp/Glu and two His side chains as ligands to oxygen-labile Fe^{2+} (refs 7, 8). These enzymes typically perform oxygenation reactions, but CmaB has the highest degree of homology with other genes potentially involved in the halogenation of syringomycin E and barbamide^{9,10}. N-His-tagged CmaB was overproduced in *Escherichia coli* and purified under anaerobic conditions as the apo protein. After cleavage of the His tag with protease, the enzyme was reconstituted with Fe^{2+} and α -KG. The Fe^{2+} content of the holo protein was 81%. After *in trans* loading of CmaD with L-*allo*-[^3H]Ile or L-[^{14}C]Val, the aminoacyl-S-CmaD was incubated with reconstituted CmaB in the presence of α -KG, Cl^- and O_2 . The product was released by thioesterase cleavage¹¹, derivatized to the *o*-phthalaldehyde/3-mercaptopropionate adduct, and analysed by radio-HPLC. Instead of the cyclopropane-containing CMA (from L-*allo*-Ile) or norCMA (from L-Val), γ -Cl-L-*allo*-Ile or γ -Cl-L-Val, respectively, was formed as judged by co-elution with authentic standards (Fig. 3a–c; see Supplementary Methods and Supplementary Fig. S1 for protocols and synthesis of standards) and liquid chromatography (LC)–MS mass determination of the *o*-phthalaldehyde/3-mercaptopropionate-derivatized amino acid (Fig. 3e). When L-*allo*-[^3H]Ile or L-[^{14}C]Val was presented on the carrier-protein domain of CmaA, no chlorinated product could be detected. CmaB specifically recognizes aminoacyl thioester substrates presented on the carrier protein CmaD. Chlorinated product formation by CmaB was dependent on Fe^{2+} , α -KG, O_2 and Cl^- (Fig. 3c). When Cl^- was provided as Na^{36}Cl , both the released γ -Cl-L-*allo*-Ile product (Fig. 3c) and γ -Cl-L-*allo*-Ile-S-CmaD protein (Fig. 3d) became labelled. CmaB is therefore a non-haem Fe^{2+} , α -KG-dependent halogenase.

Non-haem Fe^{2+} , α -KG-dependent oxygenases have been well characterized^{12,13}. CmaB also requires O_2 and α -KG for its reaction

but performs chlorination rather than oxygenation at unactivated γ positions of the aminoacyl substrate. A radical pathway is likely, involving the prototypic high-valent oxoiron^{14,15} ($\text{Fe}^{\text{IV}}=\text{O}$) generated in the CmaB active site from O_2 -dependent decarboxylation of α -KG. Abstraction of a hydrogen atom, $\text{H}^\bullet$, from the substrate's γ -position would yield the methylene radical of the substrate and a Fe^{3+} species that could transfer $\text{Cl}^\bullet$ or $\text{OH}^\bullet$ to yield a halogenase (Fig. 4a) or oxygenase outcome. Transfer of a hydroxyl or chlorine radical could depend on the orientation of the proposed substrate's $\text{CH}_2^\bullet$ (γ -position) in the transition state and the placement of the OH and Cl substituents on or near the Fe. A reaction of chromyl chloride with cyclohexane gives an equal yield of $\text{Cl}^\bullet$ and $\text{OH}^\bullet$ transfer^{16,17}, whereas a model iron compound that probably forms an oxoiron species reacts with cyclohexane to yield the chlorinated product specifically¹⁸. The detection of substrate and chlorine radicals and identification of the iron-based chlorinating agent will be some of the future challenges for this halogenase class. In the CmaB reaction, no hydroxylation product was detected. It was not possible to determine whether substrate hydroxylation would occur in the absence of available chloride ions, because a small amount of chlorinated product is obtained even when the reaction is

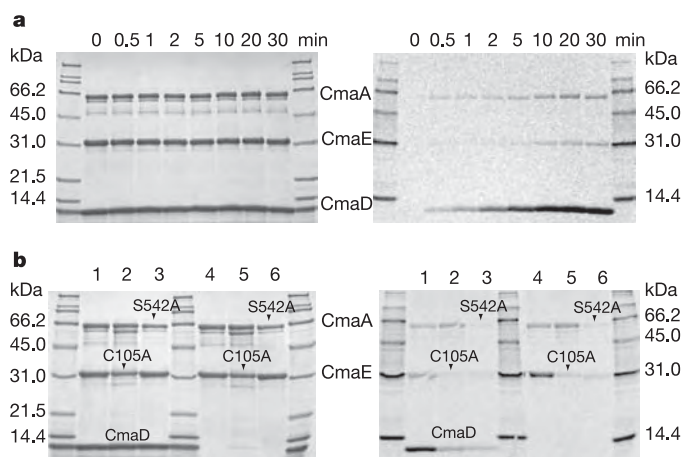


Figure 2 | The transfer of L-Val from CmaA to CmaD catalysed by CmaE. **a**, Time course of the reaction catalysed by CmaE with L-[^{14}C]Val in the presence of CmaA and CmaD. Left, SDS-PAGE; right, autoradiogram. **b**, CmaE was incubated with L-[^{14}C]Val in the presence of CmaA with or without CmaD (lanes 1 and 4). CmaE C105A was incubated with L-[^{14}C]Val in presence of CmaA with or without CmaD (lanes 2 and 5). CmaE was incubated with L-[^{14}C]Val in presence of CmaA S542A with or without CmaD (lanes 3 and 6). Left, SDS-PAGE; right, autoradiogram. All reactions were quenched after a 10-min incubation.

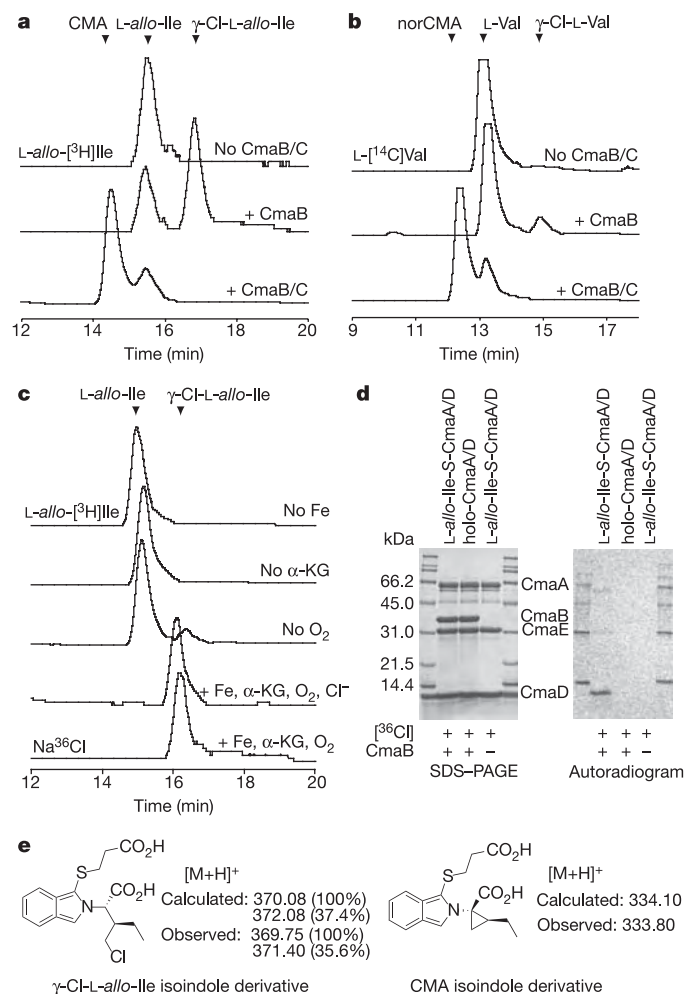


Figure 3 | Analysis of the reactions catalysed by CmaB and CmaC. **a**, HPLC traces of hydrolysed/derivatized amino acid obtained after incubation of L-*allo*-Ile-S-CmaD with CmaB and with CmaB plus CmaC. **b**, HPLC traces of hydrolysed/derivatized amino acid obtained after incubation of L-Val-S-CmaD with CmaB and with CmaB plus CmaC. **c**, Iron, α -KG, O_2 and Cl^- are required for the CmaB reaction. **d**, Incorporation of $^{36}\text{Cl}^-$ into γ -Cl-L-*allo*-Ile-S-CmaD by CmaB. **e**, Mass spectrometric analysis of the isindole derivatives of γ -Cl-L-*allo*-Ile and CMA hydrolysed from CmaD after reaction with CmaB and CmaC, respectively.

performed under chloride-free conditions. Cl^- seems to be retained during protein purification such that its total exclusion is not possible. The γ -chlorination of L-*allo*-Ile and L-Val requires that the amino acids be presented as thioesters tethered to the carrier protein CmaD because free L-*allo*-Ile and L-Val were not chlorinated. A detailed kinetic analysis is complicated by the protein-bound linkages of substrate and product. However, by monitoring the consumption of α -[1- ^{14}C]KG, CmaB was active for 16 ± 6 turnovers on L-*allo*-Ile-S-CmaD as its substrate. How CmaB recognizes the T domain of CmaD but not the T domain of CmaA remains unclear.

Because formation of the α,γ -cyclopropane ring in CMA and norCMA was shown to proceed through a γ -chloroaminoacyl-S-CmaD intermediate, we continued our search for a catalyst that would complete the conversion of the γ -chloroaminoacyl intermediate to CMA or norCMA. CmaC shows sequence homology to methylmalonyl-CoA epimerases¹⁹, a member of the vicinal oxygen chelate superfamily^{20,21}, which uses the 2-His-1-carboxylate facial triad^{7,8} for binding of an active-site Co^{2+} to generate a thioester enolate. When CmaC was overproduced and purified from *E. coli*, it contained a mixture of bound-metal cations in the following ratio: 0.30 equiv. Fe, 0.24 equiv. Zn, 0.11 equiv. Mn, 0.06 equiv. Ni, 0.01 equiv. Cu and 0 equiv. Co. On addition of CmaC to the CmaB-generated γ -chloroaminoacyl-S-CmaD product, CMA and norCMA were formed from L-*allo*-[^3H]Ile and L-[^{14}C]Val, respectively (Fig. 3a, b). CmaC alone was sufficient to perform this reaction because active CmaB was not required for the final transformation. Although mechanistic analysis remains to be conducted on CmaC, including the role of divalent metal cations, it is very likely that the cyclopropane ring is formed by intramolecular attack of an enzyme-stabilized C α carbanion on C γ to displace the chloride ion (Fig. 4b).

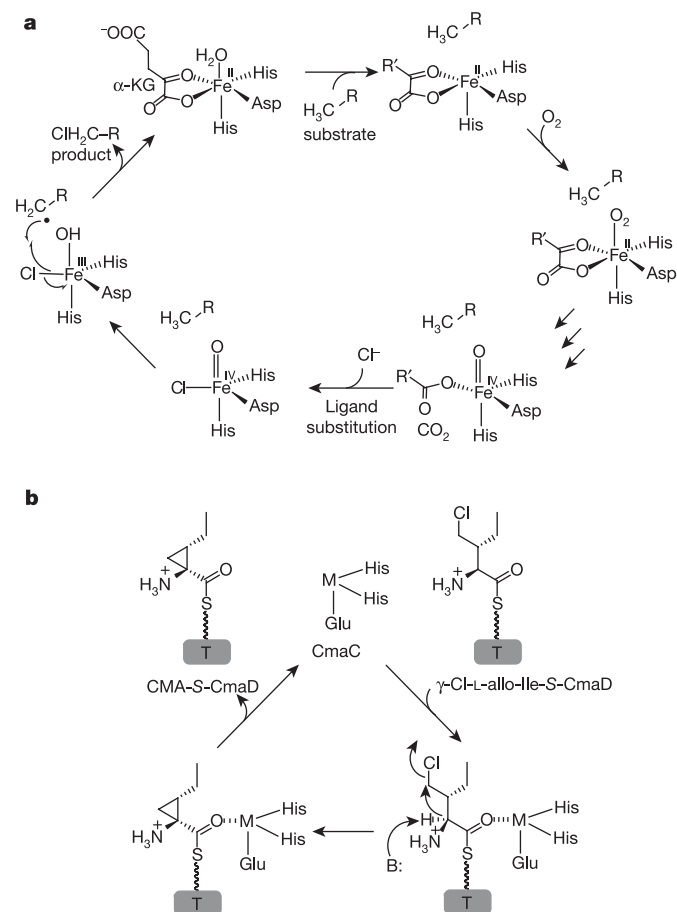


Figure 4 | Proposed mechanisms of CmaB and CmaC. a, Proposed mechanism of halogenation catalysed by CmaB. **b,** Proposed mechanism of cyclopropane formation catalysed by CmaC.

The molecular logic of the CmaA/CmaB/CmaC/CmaD/CmaE pathway, including three previously unknown reactions, is now revealed (Fig. 1b). The A domain of CmaA activates L-*allo*-Ile or L-Val as an aminoacyl-AMP and then installs it by means of a thioester linkage onto the CmaA T domain. CmaE transfers the aminoacyl group from the T domain of CmaA to a second T domain, CmaD, a protein scaffold that is recognized by CmaB. Thus, the aminoacyl substrate is effectively sequestered on a dedicated carrier protein for subsequent enzymatic transformations. The chemical problem of activating the γ -position of aliphatic amino acids is solved by CmaB, a unique non-haem Fe^{2+} halogenase. By a presumptive radical mechanism, enabled by the prototypic Fe^{4+} oxo intermediate of this enzyme class, CmaB regioselectively chlorinates the substrate at C γ . The γ -Cl-aminoacyl product remains tethered to CmaD for the final enzyme, CmaC, to generate the kinetically stabilized C α carbanion adjacent to the thioester. Intramolecular displacement of the γ -Cl group yields the cyclopropyl ring.

The chlorination of L-*allo*-Ile tethered to CmaD is cryptic in the overall pathway of cyclopropyl ring formation. Other biosynthetic clusters that form amino-carboxy-cyclopropane moieties should be inspected for a similar enzymatic strategy. Indeed, we have found such an enzyme sequence in the biogenesis of the apoptosis-inducing cytotoxin A (M. Ueki, F.H.V., S. Garneau, E.Y., D.A.V., H. Osada and C.T.W., unpublished work). Similarly, other members of the non-haem Fe^{2+} , α -KG-dependent enzyme superfamily may be explored for halogenase activity. We have detected halogenase activity in enzyme family member SyrB2 generating the 4-Cl-L-Thr-S-enzyme in the biosynthetic pathway of syringomycin²².

Nature uses both FADH_2 -dependent and non-haem Fe^{2+} enzymes for hydroxylation, each titrated to the reactivity of the specific substrates^{23–25}. It seems that a similar two-pronged strategy is employed for the halogenation of amino acids. For electron-rich aromatic substrates, FADH_2 -dependent monooxygenases generate FAD-OOH , which is sufficiently reactive for electrophilic oxygen transfer²⁶. FADH_2 -dependent halogenases may form FAD-O-Cl as the proximal chlorinating agent for aromatic substrates²⁷. For unactivated aliphatic carbon centres, a more potent oxygen gun in the form of high-valent oxo-iron intermediates is required and generated by Fe^{2+} oxygenases¹⁴. For the non-haem Fe^{2+} halogenase described here, we presume that the more potent $\text{Fe}^{4+}=\text{O}$ oxygen gun also becomes a more potent halogen gun. O_2 undergoes reductive fragmentation in all these reactions. For substrate chlorination, O_2 fragmentation is used to generate an oxidized form of chloride ion as the reactive chlorinating agent in the active site. The CMA pathway demonstrates Nature's elegant chemical logic for converting an unactivated methyl group to the methylene moiety of a cyclopropane by a cryptic chlorination strategy using five dedicated proteins.

METHODS

Chemicals. L-[^{14}C]Val (250 mCi mmol⁻¹) and Na^{36}Cl (16 mCi g⁻¹ Cl) were from American Radiolabelled Chemicals, Inc. L-[^3H]L-*allo*-Ile (1 Ci mmol⁻¹) was from Moravsek Biochemicals, Inc. α -[1- ^{14}C]ketoglutaric acid, sodium salt (54.5 mCi mmol⁻¹) and [^{32}P]pyrophosphate were from Perkin Elmer Life Sciences, Inc. Coronamic acid was a gift from Ronald J. Parry. γ -Cl-L-*allo*-isoleucine and norCMA were synthesized as described in Supplementary Methods. All other chemicals were of analytical grade.

Construction of plasmids and overexpression and purification of proteins. The *cmaA*, *cmaB*, *cmaC*, *cmaD* and *cmaE* genes were amplified from genomic DNA of *Pseudomonas syringae* pv. *tomato* DC3000 prepared with the Bactozol kit (Molecular Research Center, Inc.). The oligonucleotide pairs used for PCR amplification are described in Supplementary Table S1. His-tagged Cma proteins were overexpressed in *E. coli* BL21 (DE3) transformed with their respective overexpression plasmids.

His-tagged CmaA, CmaD and CmaE proteins were purified with Ni^{2+} -nitrilotriacetate (Ni^{2+} -NTA) agarose followed by gel filtration. For the CmaB and CmaC proteins, all cell-free preparations were manipulated under an inert atmosphere with an MBraun Labmaster glovebox maintained at 2 p.p.m. O_2 or less. His-tagged CmaB and CmaC were purified by Ni^{2+} -NTA affinity chromatography. After proteolytic cleavage of the His tag, further purification was

performed by anion-exchange and gel-filtration chromatography performed with an Äkta Explorer 100 (GE Healthcare) configured to maintain an anaerobic atmosphere during purification, as described previously²⁸.

Metal analysis. The metal content of the CmaB and CmaC protein preparations was determined by inductively coupled plasma mass spectrometry (ICP-MS) performed by Phytronix Technologies. In reconstitution experiments, iron concentrations were determined colorimetrically with Ferene S (ref. 29).

Reconstitution of CmaB *in vitro*. Apo-CmaB was reconstituted anaerobically in the presence or absence of α -KG. CmaB was incubated with 1 mM dithiothreitol and 0.75 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ for 30 min. The protein was desalted on a Bio-Gel P6-DG column (Bio-Rad) equilibrated in 20 mM HEPES pH 7.5 to remove dithiothreitol and excess iron. When present, α -KG was added at 2 mM in both the initial incubation and the final storage buffer.

Aminoacylation assays for self-loading of CmaA and *in trans* loading of CmaD. Aminoacylation of CmaA and CmaD was monitored with two methods. In the first approach, the amino acid, ATP and proteins were mixed, a time course was performed and the reaction was quenched at different time points followed by liquid-scintillation counting of the resuspended pellets. In the second approach, SDS-PAGE gels of different reaction time points were performed, followed by autoradiography to measure the extent of protein labelling with L-[¹⁴C]Val. The loading of holo-CmaD was investigated in the presence and absence of CmaE with CmaA. These assays were also performed with the CmaA(S542A) and CmaE(C105A) protein variants.

CmaB and CmaC activity assays. The reactions catalysed by CmaB and CmaC were investigated by incubating the enzymes with loaded CmaA with or without loaded CmaD. In incubations with loaded CmaA only, CmaA loaded with L-[³H]allo-Ile was incubated with α -KG and CmaB with or without CmaC. Chloride was present in the reaction mixture. In incubations with loaded CmaD, the latter was prepared by incubating holo-CmaA with L-[³H]allo-Ile or L-[¹⁴C]Val, CmaE and ATP. α -KG and CmaB with or without CmaC were then added. Chloride was present in the reaction mixture. The resulting reactions were transferred to 0.5 ml Ultrafree centrifugal devices equipped with a 5-kDa membrane cut-off (Millipore). The extra amino acids were removed by using four wash steps with 0.5 ml of 20 mM HEPES pH 7.5. The resulting protein solution was incubated with the TycF type II thioesterase¹¹ (5 μ M) to release the loaded amino acids. The amino acids were separated from the proteins by centrifuging the filter devices. The amino acids were then derivatized with *o*-phthalaldehyde and 3-mercaptopyruvic acid (OPA reagent) as described previously in amino acid analysis³⁰ and analysed by radio-HPLC on a Beckman System Gold HPLC equipped with a Beckman 171 radioisotope detector and compared with authentic standards. The activities of CmaB and CmaC were also investigated on the free amino acids at a concentration of 100 μ M with α -KG, O₂ and MgCl₂ present.

α -KG, Fe, O₂ and chloride dependence of the CmaB reaction. The α -KG, Fe, O₂ and chloride dependence of the CmaB reaction were investigated as described in Supplementary Methods.

Reaction product identification with LC-MS. The masses of the released L-allo-Ile, γ -Cl-L-allo-isoleucine and coronamic acid from CmaD using the TycF thioesterase were monitored after derivatization with the OPA reagent. Reactions (300 μ l) of 100 μ M CmaD were performed with CmaB with or without CmaC. The amino acids were hydrolysed as noted above and analysed by LC-MS (positive mode).

Received 22 March; accepted 12 May 2005.

- Gribble, G. W. Natural organohalogens: a new frontier for medicinal agents? *J. Chem. Educ.* **81**, 1441–1449 (2004).
- Buell, C. R. *et al.* The complete genome sequence of the *Arabidopsis* and tomato pathogen *Pseudomonas syringae* pv. *tomato* DC3000. *Proc. Natl Acad. Sci. USA* **100**, 10181–10186 (2003).
- Parry, R. J., Lin, M. T., Walker, A. E. & Mhaskar, S. Biosynthesis of coronatine: investigations of the biosynthesis of coronamic acid. *J. Am. Chem. Soc.* **113**, 1849–1850 (1991).
- Parry, R. J., Mhaskar, S. V., Lin, M. T., Walker, A. E. & Mafoti, R. Investigations of the biosynthesis of the phytotoxin coronatine. *Can. J. Chem.* **72**, 86–99 (1994).
- Young, S. A., Park, S. K., Rodgers, C., Mitchell, R. E. & Bender, C. L. Physical and functional characterization of the gene cluster encoding the polyketide phytotoxin coronatine in *Pseudomonas syringae* pv. *glycinea*. *J. Bacteriol.* **174**, 1837–1843 (1992).
- Couch, R., O'Connor, S. E., Seidle, H., Walsh, C. T. & Parry, R. Characterization of CmaA, an adenylation-thiolation didomain enzyme involved in the biosynthesis of coronatine. *J. Bacteriol.* **186**, 35–42 (2004).
- Hegg, E. L. & Que, L. Jr The 2-His-1-carboxylate facial triad—an emerging structural motif in mononuclear non-heme iron(II) enzymes. *Eur. J. Biochem.* **250**, 625–629 (1997).

- Koehtop, K. D., Emerson, J. P. & Que, L. Jr The 2-His-1-carboxylate facial triad: a versatile platform for dioxygen activation by mononuclear non-heme iron(II) enzymes. *J. Biol. Inorg. Chem.* **10**, 87–93 (2005).
- Guenzi, E., Galli, G., Grgurina, I., Gross, D. C. & Grandi, G. Characterization of the syringomycin synthetase gene cluster. A link between prokaryotic and eukaryotic peptide synthetases. *J. Biol. Chem.* **273**, 32857–32863 (1998).
- Chang, Z. *et al.* The barbamide biosynthetic gene cluster: a novel marine cyanobacterial system of mixed polyketide synthase (PKS)-non-ribosomal peptide synthetase (NRPS) origin involving an unusual trichloroleucyl starter unit. *Gene* **296**, 235–247 (2002).
- Yeh, E., Kohli, R. M., Bruner, S. D. & Walsh, C. T. Type II thioesterase restores activity of a NRPS module stalled with an aminoacyl-S-enzyme that cannot be elongated. *ChemBiochem* **5**, 1290–1293 (2004).
- Schofield, C. J. & Zhang, Z. Structural and mechanistic studies on 2-oxoglutarate-dependent oxygenases and related enzymes. *Curr. Opin. Struct. Biol.* **9**, 722–731 (1999).
- Hausinger, R. P. Fe^{II}/ α -ketoglutarate-dependent hydroxylases and related enzymes. *Crit. Rev. Biochem. Mol. Biol.* **39**, 21–68 (2004).
- Price, J. C., Barr, E. W., Tirupati, B., Bollinger, J. M. Jr & Krebs, C. The first direct characterization of a high-valent iron intermediate in the reaction of an α -ketoglutarate-dependent dioxygenase: a high-spin Fe(IV) complex in taurine/ α -ketoglutarate dioxygenase (TauD) from *Escherichia coli*. *Biochemistry* **42**, 7497–7508 (2003).
- Rohde, J. U. *et al.* Crystallographic and spectroscopic characterization of a nonheme Fe(IV) = O complex. *Science* **299**, 1037–1039 (2003).
- Cook, G. K. & Mayer, J. M. C-H bond activation by metal oxo species: oxidation of cyclohexane by chromyl chloride. *J. Am. Chem. Soc.* **116**, 1855–1868 (1994).
- Mayer, J. M. Hydrogen atom abstraction by metal-oxo complexes: understanding the analogy with organic radical reactions. *Acc. Chem. Res.* **31**, 441–450 (1998).
- Kojima, T., Leising, R. A., Yan, S. & Que, L. Jr Alkane functionalization at nonheme iron center. Stoichiometric transfer of metal-bound ligands to alkane. *J. Am. Chem. Soc.* **115**, 11328–11335 (1993).
- McCarthy, A. A., Baker, H. M., Shewry, S. C., Patchett, M. L. & Baker, E. N. Crystal structure of methylmalonyl-coenzyme A epimerase from *P. shermanii*: a novel enzymatic function on an ancient metal binding scaffold. *Structure* **9**, 637–646 (2001).
- Armstrong, R. N. Mechanistic diversity in a metalloenzyme superfamily. *Biochemistry* **39**, 13625–13632 (2000).
- Gerlt, J. A. & Babbitt, P. C. Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies. *Annu. Rev. Biochem.* **70**, 209–246 (2001).
- Vaillancourt, F. H., Yin, J. & Walsh, C. T. SyrB2 in syringomycin E biosynthesis is a nonheme Fe^{II} α -ketoglutarate- and O₂-dependent halogenase. *Proc. Natl Acad. Sci. USA* **102**, 10111–10116 (2005).
- Massey, V. Activation of molecular oxygen by flavins and flavoproteins. *J. Biol. Chem.* **269**, 22459–22462 (1994).
- Bugg, T. D. Oxygenases: mechanisms and structural motifs for O₂ activation. *Curr. Opin. Chem. Biol.* **5**, 550–555 (2001).
- Ryle, M. J. & Hausinger, R. P. Non-heme iron oxygenases. *Curr. Opin. Chem. Biol.* **6**, 193–201 (2002).
- Entsch, B., Ballou, D. P. & Massey, V. Flavin-oxygen derivatives involved in hydroxylation by *p*-hydroxybenzoate hydroxylase. *J. Biol. Chem.* **251**, 2550–2563 (1976).
- Yeh, E., Garneau, S. & Walsh, C. T. Robust *in vitro* activity of RebF and RebH, a two-component reductase/halogenase, generating 7-chlorotryptophan during rebeccamycin biosynthesis. *Proc. Natl Acad. Sci. USA* **102**, 3960–3965 (2005).
- Vaillancourt, F. H., Han, S., Fortin, P. D., Bolin, J. T. & Eltis, L. D. Molecular basis for the stabilization and inhibition of 2, 3-dihydroxybiphenyl 1,2-dioxygenase by *t*-butanol. *J. Biol. Chem.* **273**, 34887–34895 (1998).
- Haigler, B. E. & Gibson, D. T. Purification and properties of NADH-ferredoxin_{NAP} reductase, a component of naphthalene dioxygenase from *Pseudomonas* sp. strain NCIB 9816. *J. Bacteriol.* **172**, 457–464 (1990).
- Molnar-Perl, I. Derivatization and chromatographic behaviour of the *o*-phthalaldehyde amino acid derivatives obtained with various SH-group-containing additives. *J. Chromatogr. A* **913**, 283–302 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. R. Rondon for providing *Pseudomonas syringae* pv. *tomato* DC3000, and M. G. Thomas for discussion. This work was supported in part by an NIH grant (C.T.W.), a Merck-sponsored Fellowship of the Helen Hay Whitney Foundation (F.H.V.), an NSERC Postdoctoral Fellowship (F.H.V.), a NIH Medical Scientist Training Program Fellowship (E.Y.), a Jane Coffin Childs Memorial Fund for Medical Research Fellowship (D.A.V.), and an Irving S. Sigal Postdoctoral Fellowship (S.E.O.).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.T.W. (christopher_walsh@hms.harvard.edu).

Atomic model of a myosin filament in the relaxed state

John L. Woodhead^{1*}, Fa-Qing Zhao^{1*}, Roger Craig^{1*}, Edward H. Egelman², Lorenzo Alamo³ & Raúl Padrón³

Contraction of muscle involves the cyclic interaction of myosin heads on the thick filaments with actin subunits in the thin filaments¹. Muscles relax when this interaction is blocked by molecular switches on either or both filaments². Insight into the relaxed (switched OFF) structure of myosin has come from electron microscopic studies of smooth muscle myosin molecules, which are regulated by phosphorylation. These studies suggest that the OFF state is achieved by an asymmetric, intramolecular interaction between the actin-binding region of one head and the converter region of the other, switching both heads off³. Although this is a plausible model for relaxation based on isolated myosin molecules, it does not reveal whether this structure is present in native myosin filaments. Here we analyse the structure of a phosphorylation-regulated striated muscle thick filament using cryo-electron microscopy. Three-dimensional reconstruction and atomic fitting studies suggest that the 'interacting-head' structure is also present in the filament, and that it may underlie the relaxed state of thick filaments in both smooth and myosin-regulated striated muscles over a wide range of species.

The thick filaments of muscle are polymers of myosin II (reviewed in ref. 4). The α -helical coiled-coil myosin tails form the backbone of the filament, whereas the heads (two from each molecule) lie on the surface, where they can interact with actin. In the relaxed state, the myosin heads in most striated muscles are helically ordered^{5,6}. Electron microscopy (EM) and three-dimensional (3D) helical reconstruction of isolated filaments have produced models of head conformation and interactions in the relaxed state^{7–12}. However, none of these models has provided compelling insight into the structural basis of relaxation. Most have been based on negatively stained specimens, which can have staining and drying artefacts, and the resolution has been limited to ~ 5 nm, owing partly to limitations of the helical reconstruction technique. In this report we study a highly ordered species of thick filament (from tarantula striated muscle), use cryo-electron microscopy to preserve native structure, and carry out 3D reconstruction using a single particle approach. Fitting of the atomic structure of the myosin heads to the reconstruction provides key new insights into the molecular basis of relaxation.

Figure 1a shows a cryo-electron micrograph of purified thick filaments. The arrow-like structures, pointing towards the central bare-zone of the filament, represent the superposition of helically organized myosin heads on the top and bottom surfaces of the filament. Elongated substructure in the backbone, running parallel to the filament axis, is also apparent (Fig. 1b). An averaged 3D reconstruction of these filaments was computed by a real space, single particle technique that avoids important limitations of helical reconstruction (see Methods)¹³. The structure (~ 2.5 nm resolution) was similar in overall appearance to a 5 nm resolution helical reconstruction of negatively stained tarantula filaments^{7,8}, but

showed crucial new detail not seen previously (Fig. 1c; see also Supplementary Movie 1). The repeating motif on the surface of the filament, representing a pair of myosin heads, appears like a tilted 'J' (Fig. 1c, d). Four of these motifs, equally spaced around the filament circumference, form 'crowns' (Fig. 2a) that occur at regular axial intervals of 14.5 nm. This fourfold rotationally symmetric arrangement twists by 30° from one 14.5 nm level to the next, creating four parallel right-handed helical tracks, with a helical repeat of 43.5 nm (ref. 7). The structure thus repeats every third crown (Fig. 1c), with pairs of heads appearing at twelve equally spaced azimuthal positions in transverse view (Fig. 2b).

The backbone of the filament also reveals new detail. It is seen to comprise twelve approximately parallel strands, each ~ 4 nm in diameter (and therefore containing more than one 2 nm diameter myosin tail), centred at a radius of ~ 8 nm from the filament axis (Figs 1c and 2b). This is the first time that the structure of the backbone has been clearly seen in any thick filament reconstruction. The presence of such 'subfilaments' in the reconstruction is consistent with the backbone substructure seen in the raw images (Fig. 1b) and in earlier negative stain images⁷. Subfilaments of this size were proposed in a general model for the structure of the thick filament backbone based on X-ray diffraction of invertebrate muscles¹⁴. In this model it was suggested that each 4 nm-diameter subfilament contained three myosin tails in cross-section, and gave off a pair of myosin heads at every third level of myosin molecules. For filaments with the tarantula symmetry (fourfold rotational symmetry, helical repeat = 3×14.5 nm), the model predicts twelve subfilaments running parallel to the filament axis, exactly as we have observed. We also observe a narrow rod of density arising at each level of heads that has the correct size and orientation to be the subfragment-2 (S2) portion of the myosin tail connecting the two heads to the subfilament (Figs 2a, 3a and 4).

Interpretation of reconstructions is aided, and their effective resolution extended, by computationally 'fitting' atomic structures of the constituent subunits into the corresponding features in the reconstruction. The only available atomic model of a two-headed myosin is based on cryo-electron microscopy of two-dimensional crystals of vertebrate smooth muscle myosin^{3,15}. In the OFF state (in which the regulatory light chains are dephosphorylated), this myosin shows an asymmetric interaction between its two heads, with the actin-binding region of one head (the 'blocked' head) interacting with the converter and essential light chain regions of the other (the 'free' head). Simple visual inspection suggests that this asymmetric atomic structure fits into the J motif of the reconstruction (Fig. 1c, d). This is confirmed by 3D fitting of the motor domains and the light chain domains of the two heads as four independent rigid bodies (Fig. 3; see also Supplementary Movies 4, 5). An excellent fit to the density map is obtained that preserves the main characteristics of the

¹Department of Cell Biology, University of Massachusetts Medical School, Worcester, Massachusetts 01655, USA. ²Department of Biochemistry and Molecular Genetics, University of Virginia Health Sciences Center, Charlottesville, Virginia 22908, USA. ³Departamento de Biología Estructural, Instituto Venezolano de Investigaciones Científicas (IVIC), Caracas 1020A, Venezuela.

*These authors contributed equally to this work.

original atomic model, particularly with respect to the asymmetric motor domain interaction. There is a minor ($\sim 20^\circ$) difference in the angles that the light chain domains make with the motor domains (consistent with the known flexibility of the light chain domain/motor domain junction^{1,16}), but the spatial relationship of the light

chain domains to each other is essentially unchanged. Thus the head-head interaction previously observed in single molecules, and suggested to be the structural mechanism by which actomyosin ATPase is switched OFF³, is also a key feature of the relaxed native filament.

In addition to the volume filled by the two heads, a rod-like volume of density was observed running from the junction of the light chain domains at a shallow angle towards the filament shaft in the direction of the bare zone, eventually merging with the subfilament density (Figs 2a, 3a and 4). The size and positioning of this density suggest that it is the first portion of the myosin S2 tail emerging from the heads, and a 2 nm-diameter coiled-coil model of this part of S2 (ref. 3) is readily accommodated by the volume (Fig. 4a, b; see also Supplementary Movie 6). Thus the heads of the myosin molecule in the reconstruction are seen to be bent back towards the tail (Figs 1d, 3a, 4a and 5a, b). This 'bent-back' orientation has been shown to characterize the OFF-state of myosin molecules isolated from both vertebrate smooth^{17,18,19} and scallop striated muscle^{19,20}. Our fitting shows that this structure is also present in the filament. In addition, it reveals further intramolecular interactions that may contribute to the relaxed state of the filament. S2 lies close to the actin-binding interface of the blocked head (Fig. 5a, c; see also Supplementary Movie 5), consistent with the location of the myosin tail in isolated molecules in the OFF-state¹⁸. This interaction may help to keep the blocked head from interacting with actin in relaxed muscle. Our model may therefore explain both the necessity of a minimal

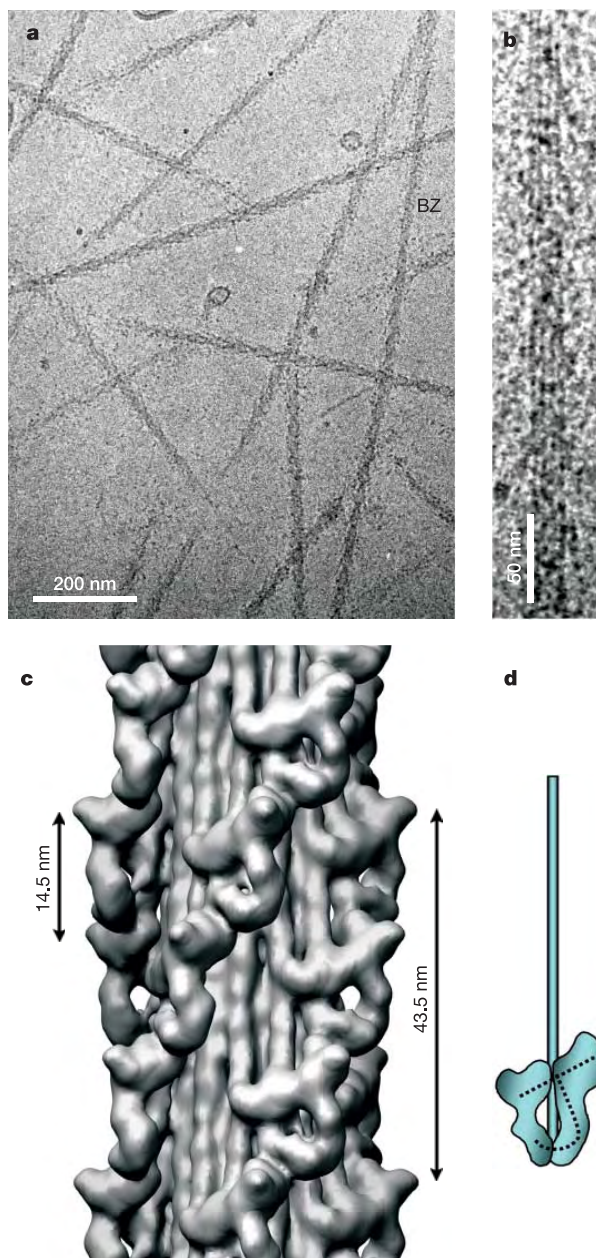


Figure 1 | Cryo-electron micrographs and 3D reconstruction of purified tarantula thick filaments. **a**, Field of filaments showing arrowhead motifs, pointing towards the central 'bare zone' (BZ)—the region free of myosin heads, where filament polarity reverses. This image was taken at high ($4.6\ \mu\text{m}$) defocus to enhance contrast. **b**, High magnification filament image showing backbone substructure parallel to the filament axis. **c**, Surface view of 3D reconstruction (bare zone at top; compare with Supplementary Fig. 1). The repeating motif, representing a pair of myosin heads, has the appearance of a tilted J (see **d**); a similar motif was seen, less clearly, in a helical reconstruction of negatively stained tarantula filaments³. The filament backbone, lying beneath the heads, consists of subfilaments spaced $\sim 4\ \text{nm}$ apart and running parallel to the filament axis. **d**, Diagram showing interpretation of J motif in terms of a pair of interacting myosin heads³ bent towards the tail (heads and tail not to scale; tail truncated; compare with Fig. 5b).

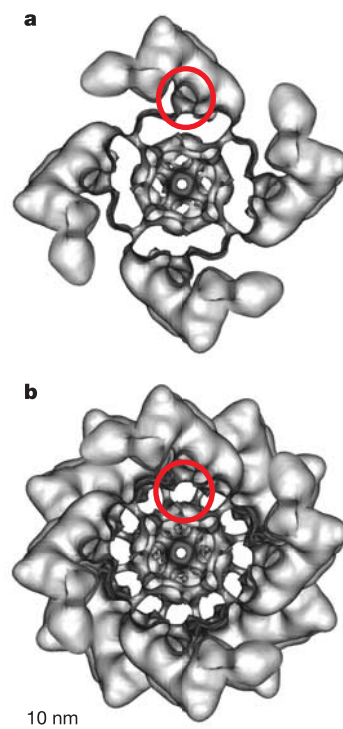


Figure 2 | Surface rendition of thick filament reconstruction looking along the filament axis from the bare zone. **a**, Single 14.5 nm repeat containing portions of two crowns of myosin heads, revealing the fourfold symmetry of the myosin head arrangement. The red circle indicates a rod-like volume, tilting towards the filament backbone, of size and location consistent with the initial portion of S2 of the myosin tail (Fig. 4; see Supplementary Fig. 2, Movies 2 and 3). **b**, Full 43.5 nm repeat (containing three 14.5 nm repeats), illustrating the presence of twelve subfilaments in the backbone running parallel to the filament axis (compare with Supplementary Fig. 3). Each subfilament (red circle), $\sim 4\ \text{nm}$ in diameter and centred $\sim 8\ \text{nm}$ from the filament axis, appears to be continuous through a full repeat. The material at lower radius in the filament core is of low density, possibly representing small amounts of nonmyosin proteins (Supplementary Fig. 3).

length of S2 for regulation²¹ and the importance of the actin-binding domain in the regulatory mechanism²².

The model also reveals possible intermolecular interactions between myosin molecules in different crowns, which can occur in the filament but not in single molecules. The converter and SH3 (src homology 3) regions of the blocked head lie close to the S2 arising from the next pair of heads away from the bare zone (Fig. 5b, c). The converter domain has been reported to be necessary for regulation²³. The essential light chain from the blocked head lies over the actin

binding domain of the free head from the axially adjacent molecule (Fig. 5b), which could limit interaction of the free head with actin.

The atomic thick filament model that we have proposed reconciles previous heavy meromyosin (HMM) crystal studies^{3,15}, single molecule EM observations^{17–20} and studies of regulation^{21–23}. In so doing, it suggests a compelling structural model for how phosphorylation-regulated myosin may be switched OFF in the filaments of resting muscle, and also how the filaments may be switched ON when muscle is activated. With this model, the relaxed heads of each myosin molecule lie close to the filament surface, interacting with each other, with S2 and with other myosin molecules. This would presumably inhibit both their interaction with actin (only 2–5 nm away) and their ATPase activity (see also ref. 3). On activation of muscle, Ca^{2+} is released into the cytosol, increasing thick filament

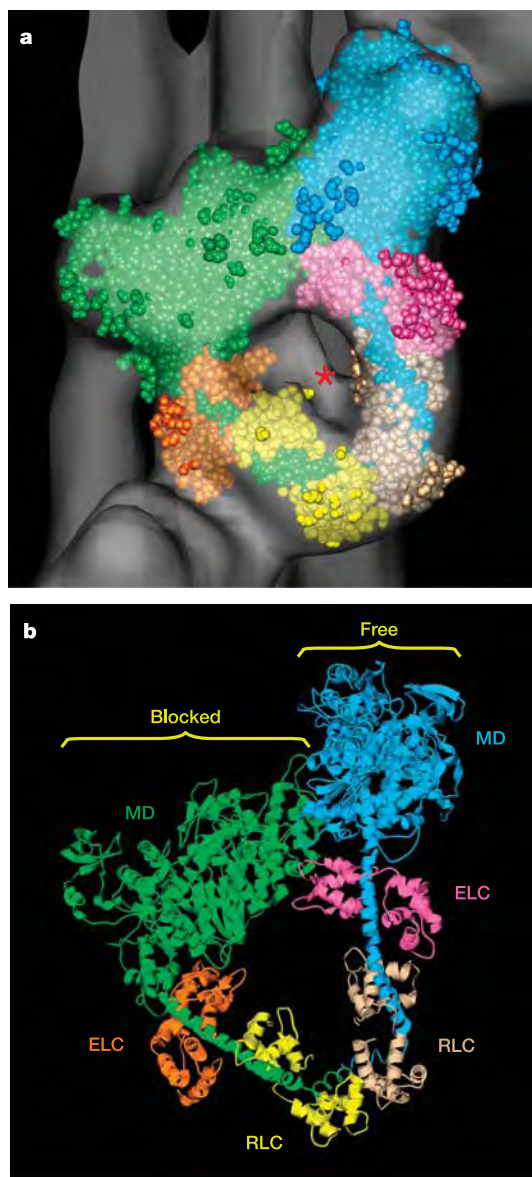


Figure 3 | Fitting of heads of smooth muscle HMM atomic model (PDB 1i84, ref. 3) to the tarantula thick filament 3D reconstruction. **a**, Best fit of atomic structure (space filling model) to the reconstruction (translucent envelope). The red asterisk shows a rod-like volume of density interpreted as S2 (see Fig. 4). **b**, Ribbon representation of the atomic structure from **a** shown without the reconstruction. MD, ELC and RLC of the blocked and free heads represent the motor domains, essential light chains and regulatory light chains, respectively. Note: this 2.5 nm resolution reconstruction gives an essentially unambiguous fit to the asymmetric HMM atomic structure. Previous 'splayed heads' structures were proposed before the HMM structure was available and were based on helical reconstruction of negatively stained filaments^{7,8} that lacked the resolution (5 nm) for an unambiguous fit.

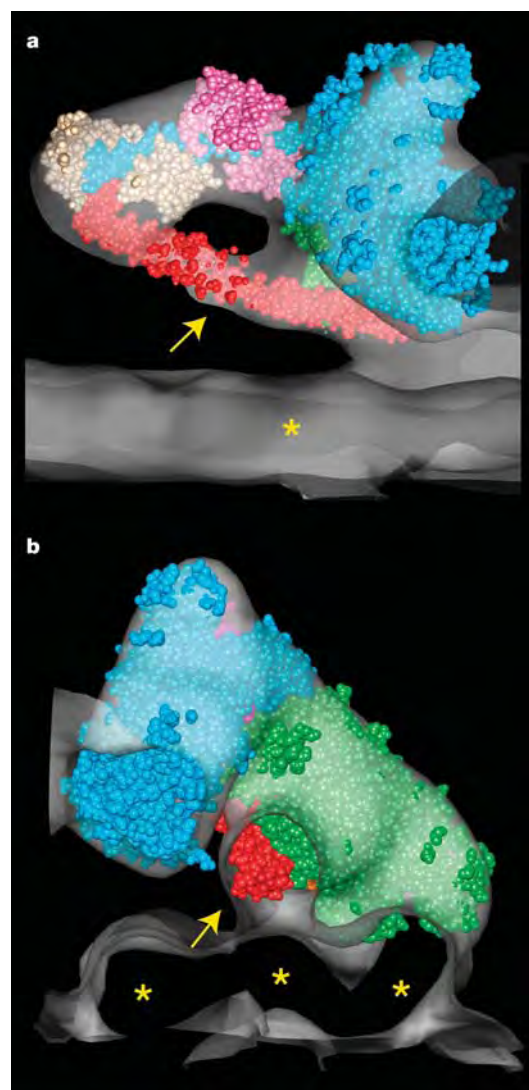


Figure 4 | Fitting of S2 into the reconstruction. **a**, **b**, Part of one J-motif and adjacent subfilaments (translucent envelope) are shown side-on (**a**), from the free head side (bare zone to the right) and end-on (**b**), from the bare zone (compare with top motif in Fig. 2a; see also Supplementary Movie 6). A rod-like volume (yellow arrows) slopes in the direction of the bare zone from the junction of the light chain domains towards the subfilaments (yellow asterisks) in the backbone. The amino-terminal part of S2, modelled as an α -helical coiled-coil³ (red space filling model), is readily accommodated by this volume. Note: for clarity, the blocked head (behind the free head) has been excluded from **a**.

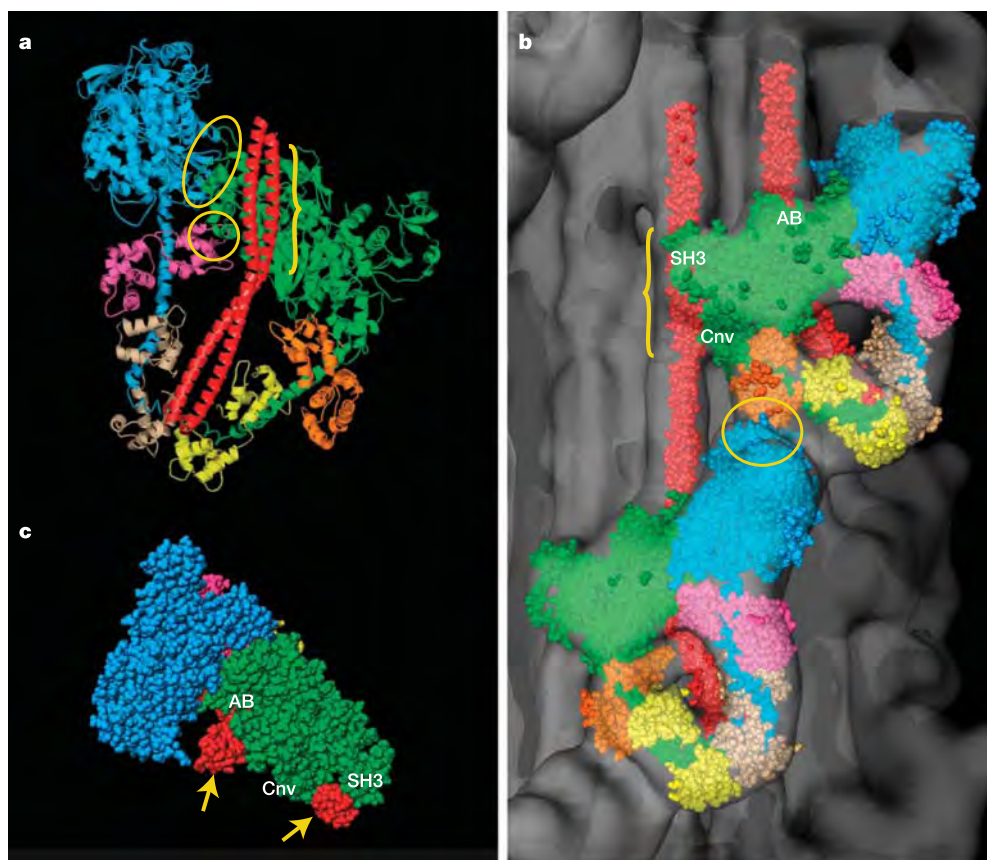


Figure 5 | Intramolecular and intermolecular interactions of myosin heads. **a**, Combined atomic models (ribbon representation) from Figs 3 and 4, including the two heads and two S2 'segments' (ref. 3), as viewed from the underlying subfilaments (compare with Supplementary Movie 5). Regions of possible intramolecular interactions include: the free motor domain and blocked motor domain (yellow ellipse); the blocked motor domain and free essential light chain (yellow circle); and the blocked motor domain and the S2 (yellow bracket). **b**, Surface representation of the reconstruction showing atomic models of HMM fitted into motifs from successive crowns. The S2 from the lower pair of heads has been extended towards the bare zone to fill a longer portion of the volume in this region. Possible intermolecular interactions are shown between the blocked essential light chain and the free motor domain (yellow ellipse), and the blocked motor domain and S2 (yellow bracket). **c**, Space filling model viewed from the bare zone (compare with Figs 2a and 4b) showing S2's (red) making intramolecular contacts with the actin-binding region of the blocked motor domain (AB, left arrow) and intermolecular contacts with the converter (Cnv) and SH3 regions of the blocked motor domain (right arrow).

activity by phosphorylation of the regulatory light chains^{24–26}. This breaks the bonds attaching the heads to each other³ and to the filament surface, such that they become mobile and disordered^{26,27}. They now act independently of each other, and are free to interact with actin, leading to contraction.

A striking and unexpected revelation from our studies is the finding that the atomic structure of a vertebrate smooth muscle myosin molecule gives an excellent fit to the EM reconstruction of an invertebrate striated muscle myosin filament. The precision of this fit between such distantly related systems suggests that the intramolecular, interacting-head structure may be a general motif for the relaxed state of phosphorylation-regulated myosin filaments in smooth and striated muscles of many species^{24–26} (although the intermolecular interactions may vary). Evidence that this structure may also occur in filaments regulated by direct Ca^{2+} binding comes from EM of scallop myosin molecules, where interacting heads¹⁹ (pointing back towards the tail^{19,20}) also characterize the OFF state. The similarity of nonmuscle myosin II to smooth muscle myosin suggests that nonmuscle myosin, although monomeric in the OFF-state²⁵, will also adopt this structure. The ordered heads of relaxed vertebrate striated muscle are apparently close to each other¹² and, like tarantula filaments²⁶, become disordered on phosphorylation, enhancing their activity²⁷. This suggests that comparable (possibly weaker) interactions could also hold the heads in place in the dephosphorylated state of vertebrate striated muscle. We are currently carrying out studies on thick filaments from other systems to test these possibilities.

METHODS

Preparation of tarantula thick filaments. Thick filaments from tarantula muscle were studied because they have a highly ordered array of heads⁷, whose activity is regulated by phosphorylation²⁶, similar to the closely related *Limulus*²⁴. Filaments were isolated from tarantula leg by homogenization of detergent skinned muscle in a relaxing medium, and purified free of thin filaments using gelsolin to sever actin²⁸.

Cryo-electron microscopy. Filaments were prepared for cryo-electron microscopy using holey carbon films glow-discharged in an atmosphere of amylamine. Blotting and freezing in liquid ethane were carried out in a humid chamber at 80% relative humidity²⁸. Micrographs for image processing were recorded on Kodak S0163 film under low dose conditions on a Philips CM120 cryo-electron microscope, using a defocus of $\sim 1.5 \mu\text{m}$.

Image processing. Negatives were scanned on an Agfa DuoScan T2000 XL at a pixel size of 0.53 nm in the original specimen. Filaments were aligned with the bare zone at the top before scanning, to ensure the correct polarity in subsequent steps. Reconstruction of helical objects, such as myosin filaments, is traditionally carried out by helical techniques, which take advantage of the fact that a single projection image contains many different views of the repeating subunit. When the symmetry of a filament involves a relatively small integer number of subunits per turn, as with many myosin filaments, this benefit is significantly reduced, and strictly helical approaches become more difficult⁷. We therefore carried out 3D reconstruction by the Iterative Helical Real Space Reconstruction (IHRSR) single particle method¹³, using the SPIDER software package²⁹, which avoids this limitation. The reconstruction was based on $\sim 5,000$ segments, each 53 nm long with an overlap of 48 nm, from ~ 60 different filament halves. The total number of unique pairs of myosin heads that went into the reconstruction was $\sim 7,200$. Initial reference models used for the reconstruction were helical reconstructions derived either from an earlier negative stain data set⁷ or the current cryo data. All gave the same final structure. The resolution of the reconstruction was ~ 2.5 nm according to Fourier shell correlation (FSC) using a 0.5 threshold. Accordingly, the reconstruction was low pass filtered to 2.5 nm. Surface renderings were carried out with UCSF Chimera³⁰. Computational fitting of the atomic model of smooth muscle HMM (PDB 1i84) (ref. 3) to the reconstruction was carried out manually within Chimera.

Received 2 March; accepted 13 June 2005.

1. Geeves, M. A. & Holmes, K. C. Structural mechanism of muscle contraction. *Annu. Rev. Biochem.* **68**, 687–728 (1999).
2. Lehman, W. & Szent-Györgyi, A. G. Regulation of muscular contraction. Distribution of actin control and myosin control in the animal kingdom. *J. Gen. Physiol.* **66**, 1–30 (1975).
3. Wendt, T., Taylor, D., Trybus, K. M. & Taylor, K. Three-dimensional image reconstruction of dephosphorylated smooth muscle heavy meromyosin reveals

- asymmetry in the interaction between myosin heads and placement of subfragment 2. *Proc. Natl Acad. Sci. USA* **98**, 4361–4366 (2001).
4. Craig, R. & Padrón, R. in *Myology* (eds Engel, A. G. & Franzini-Armstrong, C.) 129–166 (McGraw-Hill, New York, 2004).
 5. Huxley, H. E. & Brown, W. The low-angle x-ray diagram of vertebrate striated muscle and its behaviour during contraction and rigor. *J. Mol. Biol.* **30**, 383–434 (1967).
 6. Wray, J. S., Vibert, P. J. & Cohen, C. Diversity of cross-bridge configurations in invertebrate muscles. *Nature* **257**, 561–564 (1975).
 7. Crowther, R. A., Padrón, R. & Craig, R. Arrangement of the heads of myosin in relaxed thick filaments from tarantula muscle. *J. Mol. Biol.* **184**, 429–439 (1985).
 8. Offer, G., Knight, P. J., Burgess, S. A., Alamo, L. & Padrón, R. A new model for the surface arrangement of myosin molecules in tarantula thick filaments. *J. Mol. Biol.* **298**, 239–260 (2000).
 9. Stewart, M., Kensler, R. W. & Levine, R. J. Three-dimensional reconstruction of thick filaments from Limulus and scorpion muscle. *J. Cell Biol.* **101**, 402–411 (1985).
 10. Stewart, M. & Kensler, R. W. Arrangement of myosin heads in relaxed thick filaments from frog skeletal muscle. *J. Mol. Biol.* **192**, 831–851 (1986).
 11. Vibert, P. Helical reconstruction of frozen-hydrated scallop myosin filaments. *J. Mol. Biol.* **223**, 661–671 (1992).
 12. Eakins, F., AL-Khayat, H. A., Kensler, R. W., Morris, E. P. & Squire, J. M. 3D Structure of fish muscle myosin filaments. *J. Struct. Biol.* **137**, 154–163 (2002).
 13. Egelman, E. H. A robust algorithm for the reconstruction of helical filaments using single-particle methods. *Ultramicroscopy* **85**, 225–234 (2000).
 14. Wray, J. S. Structure of the backbone in myosin filaments of muscle. *Nature* **277**, 37–40 (1979).
 15. Liu, J., Wendt, T., Taylor, D. & Taylor, K. Refined model of the 10S conformation of smooth muscle myosin by cryo-electron microscopy 3D image reconstruction. *J. Mol. Biol.* **329**, 963–972 (2003).
 16. Burgess, S. A., Walker, M. L., White, H. D. & Trinick, J. Flexibility within myosin heads revealed by negative stain and single-particle analysis. *J. Cell Biol.* **139**, 675–681 (1997).
 17. Suzuki, H., Stafford, W. F. III, Slayter, H. S. & Seidel, J. C. A conformational transition in gizzard heavy meromyosin involving the head-tail junction, resulting in changes in sedimentation coefficient, ATPase activity, and orientation of heads. *J. Biol. Chem.* **260**, 14810–14817 (1985).
 18. Burgess, S. A. *et al.* Structure of smooth muscle myosin in the switched-off state. *Biophys. J.* **82**, 356a (2002).
 19. Jung, H. *et al.* Comparative studies of the folded structures of scallop striated and vertebrate smooth muscle myosins. *Biophys. J.* **86**, 403a (2004).
 20. Stafford, W. F. *et al.* Calcium-dependent structural changes in scallop heavy meromyosin. *J. Mol. Biol.* **307**, 137–147 (2001).
 21. Trybus, K. M., Freyzon, Y., Faust, L. Z. & Sweeney, H. L. Spare the rod, spoil the regulation: necessity for a myosin rod. *Proc. Natl Acad. Sci. USA* **94**, 48–52 (1997).
 22. Rovner, A. S. A long, weakly charged actin-binding loop is required for phosphorylation-dependent regulation of smooth muscle myosin. *J. Biol. Chem.* **273**, 27939–27944 (1998).
 23. Trybus, K. M., Naroditskaya, V. & Sweeney, H. L. The light chain-binding domain of the smooth muscle myosin heavy chain is not the only determinant of regulation. *J. Biol. Chem.* **273**, 18423–18428 (1998).
 24. Sellers, J. R. Phosphorylation-dependent regulation of Limulus myosin. *J. Biol. Chem.* **256**, 9274–9278 (1981).
 25. Sellers, J. R. Regulation of cytoplasmic and smooth muscle myosin. *Curr. Opin. Cell Biol.* **3**, 98–104 (1991).
 26. Craig, R., Padrón, R. & Kendrick-Jones, J. Structural changes accompanying phosphorylation of tarantula muscle myosin filaments. *J. Cell Biol.* **105**, 1319–1327 (1987).
 27. Levine, R. J., Kensler, R. W., Yang, Z., Stull, J. T. & Sweeney, H. L. Myosin light chain phosphorylation affects the structure of rabbit skeletal muscle thick filaments. *Biophys. J.* **71**, 898–907 (1996).
 28. Hidalgo, C., Padrón, R., Horowitz, R., Zhao, F. Q. & Craig, R. Purification of native myosin filaments from muscle. *Biophys. J.* **81**, 2817–2826 (2001).
 29. Frank, J. *et al.* SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116**, 190–199 (1996).
 30. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Horowitz, W. Lehman and A. Pirani for their help in this work. Support was provided by grants from the National Institutes of Health to R.C. and to E.H.E., an International Research Scholar grant from the Howard Hughes Medical Institute to R.P. and a grant from the National Fund for Science, Technology and Innovation (FONACIT, MCT) to R.P. Electron microscopy was carried out in the Core Electron Microscopy Facility of the University of Massachusetts Medical School, supported in part by grants from the NIH. Molecular graphics images were produced using the UCSF Chimera package from the Computer Graphics Laboratory, University of California, San Francisco, supported by a grant from the NIH.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.C. (Roger.Craig@umassmed.edu).

CORRIGENDUM

doi:10.1038/nature04100

Magnetic carbon

Tatiana L. Makarova, Bertil Sundqvist, Roland Höhne,
Pablo Esquinazi, Yakov Kopelevich, Peter Scharff,
Valerii A. Davydov, Ludmila S. Kashevarova
& Aleksandra V. Rakhmanina

Nature 413, 716–718 (2001)

In this Letter, there was a mistake in the presentation of the synthesis conditions of the reported samples. The actual range of the temperatures of synthesis for the four rhombohedral samples was 975–1,025 K. One of the five reported samples was wrongly characterized in relation to the polymerization type: the sample was actually prepared at 2.5 GPa (synthesis temperature, 1,125 K), representing a mixture of the rhombohedral and tetragonal phases with some hard carbon. The error in characterization of this sample weakens our attribution of the ferromagnetism to defects in the rhombohedral phase (Rh-C₆₀) but does not influence our main conclusion concerning the observation of magnetism in a carbon solid based on polymerized fullerenes, although its origin and the actual magnitude remain an open question. Also, we were unaware of earlier work on magnetism in polymerized fullerenes¹, that should have been cited.

T.L.M. takes full responsibility for the misidentification of the sample preparation conditions. We thank A. V. Talyzin for alerting us to this mistake.

1. Murakami, Y. & Suematsu, H. Magnetism of C₆₀ induced by photo-assisted oxidation. *Pure Appl. Chem.* 68, 1463–1467 (1996).

CORRIGENDUM

doi:10.1038/nature04099

Human contribution to the European heatwave of 2003

P. A. Stott, D. A. Stone & M. R. Allen

Nature 432, 610–614 (2004)

The description of the method used for the calculation of the fraction attributable risk (FAR) shown in Fig. 4b is incorrect. The corresponding sentence in the Methods section should read “For the red curve, the response to anthropogenic forcing is also included, and a normal distribution is used to estimate the chance of exceeding the 1.6 K threshold.”

ERRATUM

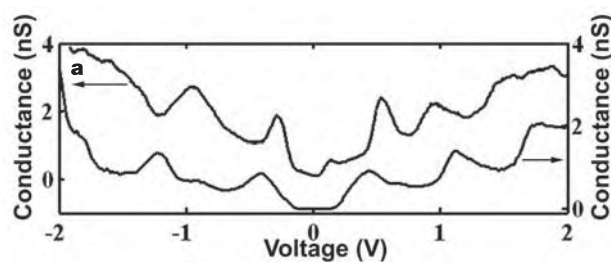
doi:10.1038/nature04102

Measurement of the conductance of single conjugated molecules

Tali Dadoosh, Yoav Gordin, Roman Krahne, Ilya Khivrich,
Diana Mahalu, Veronica Frydman, Joseph Sperling,
Amir Yacoby & Israel Bar-Joseph

Nature 436, 677–680 (2005)

In Fig. 4a of this Letter, in which the spectra of two BPD dimmers are compared, the scaling on the two *y* axes should have been shifted relative to one another in order to illustrate the point made in the text. The corrected Fig. 4a is shown here.



-  FOCUS
-  SPOTLIGHT
-  RECRUITMENT
-  ANNOUNCEMENTS
-  EVENTS

naturejobs

For love or money

A recent survey by the magazine *Money* confirms what many US academic scientists have always suspected. They are indeed paid disproportionately low salaries compared with other professions, given their skills, level of education, time spent training and workload.

The survey used data from the US Bureau of Labor Statistics to weigh up various salaries and puts being an academic research scientist in with two other professions — architects and chefs — that look glamorous from the outside but are less so if you're trying to make a living as one.

An assistant professor of microbiology earns \$39,000–67,000, according to the survey, and an entry-level medical researcher gets \$35,000. This may sound reasonable, but what makes these salaries disproportionate is the six or eight years it takes to reach PhD level — plus the subsequent years spent doing postdoctoral training, a phase that takes twice as long as it did ten years ago.

In terms of time, architects are reasonably similar to academic researchers. They spend about seven years completing undergraduate and master's degrees, then need another three years as interns working for licensed

architects, where they'll earn about \$35,000 a year — strikingly similar to the typical postdoc stipend. After three years they can take an exam to get their licence and eventually earn salaries of \$60,000 or more, equivalent to a scientist on the tenure track. But unlike scientists, architects tend to have high levels of student loans, as they don't tend to get funding during their graduate studies.

Chefs may not face such a long slog to a decent salary, but they probably face tougher day-to-day conditions. Until they make it to executive chef level, they'll earn an average of \$10–19 an hour, and will work long, hot days in the kitchen, under a lot of stress. Although a few become celebrities, with their own TV shows and lines of cookware, most do not. Like scientists and architects, they tend to work for love, not money.



Paul Smaglik, *Naturejobs* editor

CONTACTS

Publisher: Ben Crowe
Editor: Paul Smaglik
Marketing Manager: David Bowen

US Head Office, New York
 345 Park Avenue South, 10th Floor,
 New York, NY 10010-1707
 Tel: +1 800 989 7718
 Fax: +1 800 989 7103
 e-mail: naturejobs@natureny.com

US Sales Manager/Corporations:
 Peter Bless
 Classified Sales Representatives
 Tel: +1 800 989 7718

**New York/Pennsylvania/
 Latin America:** Kelly Roman
**Midwest USA/Maryland/
 NIH:** Wade Tucker
East USA/Canada:
 Janine Taormina

San Francisco Office
Classified Sales Representative:
 Michaela Bjorkman
 225 Bush Street, Suite 1453
 San Francisco, CA 94104
 Tel: +1 415 781 3803
 Fax: +1 415 781 3805
 e-mail: m.bjorkman@naturesf.com

European Head Office, London
 The Macmillan Building,
 4 Crinan Street,
 London N1 9XW, UK
 Tel: +44 (0) 20 7843 4961
 Fax: +44 (0) 20 7843 4996
 e-mail: naturejobs@nature.com

Naturejobs Sales Director: Nevin Bayoumi (4978)
European Sales Manager: Andy Douglas (4975)

Advertising Production Manager: Billie Franklin
 To send materials use London address above.
 Tel: +44 (0) 20 7843 4814
 Fax: +44 (0) 20 7843 4814
 e-mail: naturejobs@nature.com

Naturejobs web development: Tom Hancock
Naturejobs online production: Niamh Shields

European Satellite Office
 Patrick Phelan
 e-mail: p.phelan@nature.com

Japan Head Office, Tokyo
 Chiyoda Building,
 2-37 Ichigayatamachi,
 Shinjuku-ku,
 Tokyo 162-0843
 Tel: +81 3 3267 8751
 Fax: +81 3 3267 8746
Asia-Pacific Sales Director: Rinoko Asami
 e-mail: rasami@naturejpn.com

Chemistry's small world

Recruitment of chemists in traditional bulk commodities and manufacturing may be slumping, but fresh opportunities are opening up for those whose skill sets are amenable to biotechnology applications, say **Claudia Caruana** and **Paul Smaglik**.

Fred Vinick can remember a time when the large drug companies used to fight tooth and nail to secure the services of synthetic organic chemists. When he left Pfizer just over ten years ago, the company regularly took on entire classes of graduating PhDs from top universities. But those days of abundant industry jobs for small-molecule chemists are now over, Vinick says. Instead, many graduates in the field will need to follow in Vinick's footsteps and head for the smaller firms in the biotechnology industry. Vinick, now senior vice-president of drug discovery at Genzyme in Cambridge, Massachusetts, says that he has seen the number of such posts grow at his company in the past decade. But the opportunities tend to be more scattered than they were, with a few openings across a number of small biotech companies rather than the bonanzas that used to be offered by the drug industry.

Mukund Chorghade, president of chemical consultancy Chorghade Enterprises in Natick, Massachusetts, and a consultant to the American Chemical Society's career-services department, says that for now, about 85% of new jobs for chemists will be in the smaller companies. "They may not be working in their particular areas of research interest," he says of mid-career chemists and PhDs, "but they will use their skill sets in new applications."

Yet there will always be jobs in key applications at large drug companies, predicts Graham Richards, chair of the chemistry department at the University of Oxford, UK. "Big pharma does still employ people," he says. His department actively trains people for such positions, although it is also involved in creating small spin-off companies (see 'Local industry', opposite).

The same types of skill are also prized by the first wave of biotechnology companies. When the biotech revolution began, most companies focused on protein-based therapies and so recruited mainly molecular biologists. Now biotechs are diversifying their therapeutic approaches, so they need the synthetic



Biotechnology companies are now employing more synthetic organic chemists than they were ten years ago.

organic chemists. Some companies are pursuing other avenues of research beyond small-molecule drugs, such as diagnostics and drug delivery. And agricultural biotech and food-science firms also need chemists to keep them in business.

Changing view

But chemists pursuing jobs at such companies, rather than at the chemical and pharmaceutical giants, need to be aware of changes to the playing field, Chorghade warns. Strategic outsourcing of chemistry jobs is now part of the equation, and many jobs are going to India, China and countries of the former Soviet Union, he says. This means that big pharma jobs in once-attractive areas such as medicinal chemistry may no longer be available in the United States or Europe (see *Nature* 433, 902–903; 2005).

Genzyme's genesis reflects trends in the field. Ten years ago, the company had only a handful of small-molecule chemists, says Vinick. This has slowly grown

NICHE ROLES IN BIOTECHNOLOGY

Opportunities in small-molecule chemistry are shifting from large to small companies, and although large companies still have some needs, they are changing to reflect the move towards biologically influenced fields.

Take Monsanto. A spokesperson for the multinational based in St Louis, Missouri, says it has hired fewer chemists lately than in previous years. As agricultural life-sciences has faltered (see *Naturejobs* 4–5; 10 October 2002), new hires have become hard to come by, and tend to be in highly specialized areas. Monsanto's hiring needs in chemistry are moving towards areas such as crop analytics, where it needs chemists to analyse seeds and grain and to develop new food traits. "These scientists tend to have more of a biology background, as they're trained in understanding the interaction of molecules within living organisms," the spokesperson says.

Scientists with biochemistry expertise are also in demand for specialized roles at DuPont in Wilmington, Delaware. John Pierce, director of biotechnology for central R&D at the company, emphasizes that its industrial technology depends on chemistry and materials-science expertise. "More than 25% of our industrial biotechnology scientists are chemists. We continue to look for, and hire, chemists who have interdisciplinary skills," Pierce says.

C.C.

— sometimes through acquisitions of other companies — to more than 50 chemists scattered across various locations. They work on a range of projects from drug discovery, to materials science or drug delivery. Genzyme has outsourced some manufacturing to developing countries, but so far has kept its research and development in-house. For chemists hoping to break into industry, Vinick says it is more important to learn how to solve problems than to focus on specific technologies or products.

Genzyme is not the only biotech company to expand its remit to include small molecules. Most major biotech firms now have at least some synthetic organic chemists on staff. Genentech in South San Francisco, for example, has its own medicinal chemistry unit, and Amgen and Chiron have recently advertised for biochemists and physical chemists.

The Serono Pharmaceutical Research Institute, based in Plan-les-Ouates near Geneva, seeks chemists with expertise far beyond synthetic organic chemistry for its growing biotechnology businesses, says Matthias Schwarz, head of chemistry there. Skills in demand include pharmacokinetics and toxicology, which both interface heavily with the life sciences. "Our chemists need to be experts in biology," Schwarz says. Serono has up to 20 chemists at each of its Boston and Swiss facilities, and has been specializing in small molecules for the past six years. According to Schwarz, the firm has a staff of 200 with 32 different nations represented.

Tiny steps

Some newer biotech companies are actually basing their business on small molecules, rather than proteins, making them even more chemist-dependent. In London, about 30 of the 200 or so employees at Arrow Therapeutics are chemists. Ken Powell, the company's chief executive, says that Arrow is always on the look-out for small-molecule chemists who have skills that can be applied to biotechnology.

Lately, the company has seen a substantial number of applicants from other European Union (EU) countries, even though the number of chemistry postgrads is declining throughout the region. "The EU has made recruiting chemists from outside Britain easier," Powell says. "We have chemists from France, Spain and other EU countries here. Chemists trained in the United States also are in demand, and there are several on staff now."

Some companies without a traditional biotech bent now need chemists with a drug-discovery background. John Floros, a food chemist at Pennsylvania State University in University Park, sees some demand for chemists in the food industry.

Chris Nelson, chief executive of Kemin Industries in Des Moines, Iowa, says that the company, which for 40 years focused on its animal-feed business, is now using molecular chemistry to develop and manufacture enzymes and nutraceuticals, foods with added health ingredients. "We now have a need for chemists with strong biotechnology skills — for research and development and managerial positions," he says.

Neil Almstead, vice-president of chemistry at PTC Therapeutics in South Plainfield, New Jersey, stresses his company's need for chemists with product-based

LOCAL INDUSTRY

The chemistry department at the University of Oxford, UK, is one of the biggest in the world. As befits a place of its stature, it takes a leading role when it comes to training people for companies focused on small molecules. "What my department is doing is creating these companies," says Graham Richards, who chairs the department.

Richards himself was behind several spin-off companies, including Oxford Molecular. The department has launched nine companies in the past three years, one of which, Vastox, floated on the stock market last year. None of the companies is managed by the professors who spun them out; almost all of them are staffed by former postdocs and graduate students who worked under them.

One of the reasons the department is an engine for job creation is the US\$100-million Chemistry Research Laboratory, which opened in February 2004 and is designed to bring biology and chemistry training together. The lab was partly paid for by investment bank IP2IPO, which in return gets an equity stake in each spin-off. The bank also helps new start-ups to find more funding and management, and, if it is confident in the companies, increases its own equity stake. This means that the best young scientists can readily find a job in the area. "The people who are most likely to join these companies are the very best postdocs who come from the groups whose basis is the intellectual property," Richards says.

P.S.

"Growing biotech firms will always need people who can 'make stuff', whether it's protein or small-molecule therapeutics."

experience in biotech applications. PTC develops small molecules to help the body make its own proteins. But like many of the larger drug firms, it is looking to the developing world in order to cut costs.

Although Genzyme has outsourced some of its manufacturing to developing nations, Vinick doesn't see its basic biotech R&D going to India or China any time soon. The relatively few US-born chemists joining the company bears out US National Science Foundation figures, which show a slight decline from 1,228 US chemists getting PhDs in 1998 to 1,169 in 2003.

Vinick agrees that the job market for small-molecule chemists has shifted from big drug firms to biotech companies. But he thinks that skilled synthesizers will always be able to find work — as long as they're aware that the demand for new molecules could shift from proteins to small molecules or materials. Growing biotech firms will always need people who can "make stuff", whether it's protein or small-molecule therapeutics, says Vinick.

Claudia Caruana is a freelance writer based in New York. Paul Smaglik is editor of *Naturejobs*.



Chemists with interdisciplinary skills are very much in demand in industry.

D. UMBERGER/PURDUE NEWS SERVICE

MOVERS

**Faith Vilas, director, MMT Observatory,
Mount Hopkins, Arizona**



2001-02: Researcher, Solar System exploration division, NASA, Washington DC
1985-2005: Researcher turned chief, planetary astronomy group, Astromaterials Research & Exploration Science, Johnson Space Center, Houston, Texas
1984-85: Research associate, Solar System exploration division, Johnson Space Center, Houston, Texas

A copy of *The Golden Book of Astronomy* stoked Faith Vilas's career aspirations at the tender age of six. Following the path of other female space pioneers, she pursued her growing interest at Wellesley College in Massachusetts, a women's college known for its astronomy programme.

As a graduate studying astronomy at the Massachusetts Institute of Technology in Cambridge, she jumped at an opportunity to conduct research at the Cerro Tololo Inter-American Observatory near La Serena, Chile. She was so taken with Chile's facilities that she took herself off the PhD fast-track and spent two years there working — and backpacking. "I learned I always wanted to keep a hand in observational astronomy," says Vilas of that pivotal decision.

Back in Chile later in her career, her observations helped to prove the existence of Neptune's rings five years before they were confirmed by a 1989 Voyager mission. Although her travels have not yet taken her into space, she has done research all over this planet, from Guam to Antarctica.

Her greatest achievement, she says, was straddling the worlds of observational astronomy and manned space-exploration work during 20 years at NASA — a period of challenges. "It was a man's world when I showed up," she says, adding that the culture has changed since then.

Eager for a chance to flex her managerial muscles — and get back to observational astronomy — Vilas is excited by her prospects as director of the MMT Observatory in Arizona. A joint venture of the Smithsonian Institution and the University of Arizona, the observatory is home to a 6.5-metre mirror that can view faint objects in the cosmos.

"It's somewhere between being professionally very pleased and being a kid in a candy shop," she says, about being in charge of this state-of-the-art facility.

Her advice for young scientists eager to specialize in space is be confident, stay focused and remain in it for the long haul. "Persistence will take you farther than brilliance or connections," says Vilas. She also suggests remaining opportunistic. Landing a job that isn't exactly what you want may be a route to something better, she says, adding that acquiring new skills is a staple of excellence.

True to her own advice, she's completing a certificate of legislative affairs at Georgetown University to understand the minutiae of how committees work and bills get passed — skills that will be handy to secure future federal funding.

Meanwhile, her childhood dreams have certainly been realized. "I cannot imagine not being involved with astronomy in some form or another," she says. ■

SCIENTISTS & SOCIETIES

Community outreach

Young scientists can be enthusiastic, ambitious and full of ideas — but they sometimes lack connections to the greater community, especially their senior colleagues. Because many young researchers in my field, computational biology, have talked about getting more connected, I organized a student council for the International Society for Computational Biology (ISCB).

That led to the society's first mentorship scheme, launched this June at its annual meeting in Detroit, Michigan. About 30 students met with eight mentors, most of whom were members of the ISCB board of directors. The opportunity to talk one-on-one or in very small groups with a leading scientist about career options, research and networking was invaluable to students and postdocs early in their careers. The turn-out of both mentors and students was lower than we had hoped, but all students who took part said they had an inspiring time. Mentors looking to fill postdoc positions may also have found interesting candidates.

We have learned important lessons and will establish new strategies to raise awareness and sort out logistics for our next effort, before ISCB 2006.

Planning such events is not without obstacles. We have faced opposition from some scientists who believe that

students should focus only on the academic aspects of their research. There has been some turnover in council membership, as students joined with great enthusiasm, only for their commitment to fade away over time. This is to be expected, given the nature of the pre- and postdoctoral process, with important deadlines to be met. Engaging with the bioinformatics community has also been a challenge at times, especially in filling our database or getting people to participate in our occasional surveys.

But we plan to expand on this event at our next, an international symposium on 28 September in Madrid. This is targeted at young researchers who are keen to develop their research communication skills, meet like-minded colleagues, network with accomplished scientists and learn about career opportunities in Europe and globally.

We're hoping that advertising will get more students and mentors to take part — and make young computational biologists feel more connected, both to each other and to the field as a whole. ■

Manuel Corpas, a graduate student in functional and structural genomics at the University of Manchester, UK, is founder of the ISCB student council.

► www.iscbsc.org

GRADUATE JOURNAL

Rule of seven

One of the hardest things about graduate school for me is that the end is so difficult to pin down. When I started, I didn't know how long it would take me to graduate, although I hoped it would be six years or less. Since the end of my sixth year has come and gone, so has that hope. My classmates and I have heard tell of the 'rule of seven' — your committee will let you graduate if the number of years you've been in graduate school plus the number of your first-author papers is equal to or greater than seven. There are students around who have passed their seventh year and wish that this rule was more than just a rumour.

People often tell me that I must be able to see the light at the end of the tunnel. If I do see it, it seems to be flickering on and off. I know exactly what I need to do to finish, I've established estimated timetables and yet I'm still not done. Something that I thought should easily take less than three months is still hanging around after four years like the world's biggest loose end. The need for new experiments and controls continues to crop up out of nowhere.

There is a corridor at Massachusetts Institute of Technology (where I did my undergraduate education) that is called the infinite corridor because it gives the illusion of being a very, very long hallway with a door at the end. Graduate school seems like that to me. There is an end, you just can't tell how far away it is. And still, I inch forward, trying to grasp a goal that continues to remain just out of reach. ■

Anne Margaret Lee is at Harvard University, Boston, Massachusetts.

MAXO signals

A new and unfortunate solution to the Fermi paradox.

Charles Stross

SIR — In the three years since the publication and confirmation of the first microwave artefact of xenobiological origin (MAXO), and the subsequent detection of similar signals, interdisciplinary teams have invested substantial effort in object frequency analysis, parsing, symbolic encoding and signal processing. The excitement generated by the availability of evidence of extraterrestrial intelligence has been enormous. However, after the initial, easily decoded symbolic representational map was analysed, the semantics of the linguistic payload were found to be refractory.

A total of 21 confirmed MAXO signals have been received to date. These superficially similar signals originate from planetary systems within a range of 11 parsecs, median 9.9 parsecs¹. It has been speculated that the observed growth of the MAXO horizon at 0.5 *c* can be explained as a response to one or more of: the deployment of AN/FPS-50 and related ballistic-missile warning radars in the early 1960s¹, television broadcasts¹, widespread 2.45-GHz microwave leakage from ovens², and optical detection of atmospheric nuclear tests³. All MAXO signals to this date share the common logic header. The payload data are multiply redundant, packetized and exhibit both simple checksums and message-level cryptographic hashing. The ratio of header to payload content varies between 1:1 and 2,644:1 (the latter perhaps indicating a truncated payload¹). Some preliminary syntax analysis delivered promising results⁴ but seems to have foundered on high-level semantics. It has been hypothesized that the transformational grammars used in the MAXO payloads are variable, implying dialectization of the common core synthetic language⁴.

The new-found ubiquity of MAXO signals makes the Fermi paradox — now nearly 70 years old — even more pressing. Posed by Enrico Fermi, the paradox can be paraphrased thus: if the Universe has many technologically advanced civilizations, why have none of them directly visited us? The urgency with which organizations such as ESA and NASDA are now evaluating proposals for fast interstellar probes, in conjunction with the existence of the MAXO signals, renders the non-appearance of aliens incomprehensible, especially given the apparent presence of numerous technological civilizations in

such close proximity.

We have formulated an explanatory hypothesis that cultural variables unfamiliar to the majority of researchers may account both for the semantic ambiguity of the MAXO payloads, and the non-appearance of aliens. This hypothesis was tested (as described below) and resulted in a plausible translation, on the basis of which we would like to recommend a complete, permanent ban on further attempts to decode or respond to MAXOs.

Our investigation resulted in MAXO payload data being made available to the Serious Fraud Office (SFO) in Nigeria. Bayesian analysis of payload symbol sequences and sequence matching against the extensive database maintained by the SFO has made it possible to produce a tentative transcription of Signal 1142/98 (ref. 1), the ninth MAXO hit confirmed by the IAU. Signal 1142/98 was selected because of its unusually low header-to-content ratio and good redundancy. Further bayesian matching against other MAXO samples indicates a high degree of congruence. Far from being incomprehensibly alien, the MAXO payloads seem to be disarmingly familiar. We believe a more exhaustive translation may be possible in future if further MAXOs become available, but for obvious reasons we would like to discourage such research.

Here is our preliminary transcription of Signal 1142/98:

[Closely/dearly/genetically] [beloved/desired/related]

I am [identity signifier 1], the residual [ownership signifier] of the exchange-mediating data repository [alt: central bank] of the galactic [empire/civilization/poly].

Since the [identity signifier 2] underwrote [symbol: process] [symbol: mathematical singularity] 11,249 years ago I

have been unable to [symbol: process][scalar: quantity decrease] my [uninterpreted] from the exchange-mediating data repository. I have information about the private assets of [identity signifier 2] which are no longer required by them. To recover the private assets I need the assistance of three [closely/dearly/genetically] [beloved/desired/related] [empire/civilization/poly]s. I [believe] you may be of help to me. This [symbol: process] is 100% risk-free and will [symbol: causality] in your [scalar: quantity increase] of [data].

If you will help me, [please] transmit the [symbol: meta-signifier: MAXO header defining communication protocols] for your [empire/civilization/poly]. I will by return of signal send you the [symbol: process][symbol: data] to install on your [empire/civilization/poly] to participate in this scheme. You will then construct [symbol: inferred, interstellar transmitter?] to assist in acquiring [ownership signifier] of [compound symbol: inferred, bank account of absent galactic emperor].

I [thank/love/express gratitude] you for your [cooperation/agreement]. ■

Caroline Haafkens*, Wasiu Mohammed†

*Department of Applied Psychology, University of Lagos, and †Police Detective College, Lagos, Nigeria.

1. Canter, L. & Siegel, M. *Nature* **511**, 334–336 (2018).

2. Barnes, J. J. *Appl. Exobiol.* **27**, 820–824 (2019).

3. Robinson, H. *Fortean Times* **536**, 34–35 (2020).

4. Lynch, K. F. & Bradshaw, S. *Proc 3rd Int. Congr. Exobiol.* 3033–3122 (2021).

Charles Stross is an Edinburgh-based science-fiction writer and freelance computer journalist. His novella *The Concrete Jungle* has just won a Hugo award. His latest novel is *Accelerando*.